Table of Contents

Contract No: 53-3198-3-038
MPR Reference No: 8156-029

ERROR ANALYSIS OF MATH-CPS MODEL ESTIMATES

December 15, 1996

Authors:
Sally Thurston
Department of Statistics, Harvard University

Alan Zaslavsky
Department of Health Care Policy, Harvard Medical School

Submitted to:
U.S. Department of Agriculture
Food and Consumer Service
3101 Park Center Drive
2nd floor
Alexandria, VA 22302

Project Officer:
Alana Landey

Submitted by:
Datametrics Research Inc.
215 Waban Ave.
Waban, MA 02168

Under Subcontract to:
Mathematica Policy Research, Inc.
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
(202)484-9220

Project Director:
Carole Trippe

This work was prepared as one task of a competitively awarded contract; the total amount of the contract is $4,225,801.

Table of Contents

Contract No: 53-3198-3-038
MPR Reference No: 8156-029

ERROR ANALYSIS OF MATH-CPS MODEL ESTIMATES

December 15, 1996

Authors:
Sally Thurston
Department of Statistics, Harvard University

Alan Zaslavsky
Department of Health Care Policy, Harvard Medical School

Submitted to:
U.S. Department of Agriculture
Food and Consumer Service
3101 Park Center Drive
2nd floor
Alexandria, VA 22302

Project Officer:
Alana Landey

Submitted by:
Datametrics Research Inc.
215 Waban Ave.
Waban, MA 02168

Under Subcontract to:
Mathematica Policy Research, Inc.
600 Maryland Avenue, SW
Suite 550
Washington, DC 20024-2512
(202)484-9220

Project Director:
Carole Trippe

This work was prepared as one task of a competitively awarded contract; the total amount of the contract is $4,225,801.

Table of Contents

The authors thank John DiCarlo, John Czajka, and Julie Sykes for their assistance on this work. In particular, the authors gratefully acknowledge enormous assistance from John DiCarlo who helped us to understand and use the MATH-CPS model code, provided us with code for the reforms, closely followed our research, and helped us to stay focused on issues that are important to the ultimate users of the MATH-CPS model. The authors also thank Don Rubin for facilitating the development of this project. Any errors in this report are the responsibility of the authors.

Contents

List of Figures iv

List of Tables iv

Executive Summary x

1 Introduction 1

2 Sources of uncertainty in microsimulation models 3

3 Approaches and results of error analysis in other studies 7

3.1 External validation studies of the MATH-CPS model 8

3.2 Studies using sensitivity analyses 9

3.3 Studies addressing stochastic simulation, calibration, or other sources of uncertainty 11

4 Our treatment of uncertainties 15

5 The MATH-CPS model 19

5.1 Creation of the MATH-CPS database from the Current Population Survey 19

5.2 MATH-CPS model modules 21

6 Sources of uncertainty in the MATH-CPS model and our treatment of these uncertainties 25

6.1 Stochastic simulation 25

6.2 Sampling 26

6.3 Model parameters 27

6.4 Model specification 28

Table of Contents

6.5 Calibration 29

6.6 Macro effects 34

6.7 Food stamp unit size 36

7 A method of quantifying uncertainties in the MATH-CPS model 39

7.1 Simulation experiment design 39

7.2 A model for the influence of observations 42

7.3 Estimation of quantities of interest 45

8 Results from the MATH-CPS model 51

8.1 Effect of number of modules on stochastic variability 51

8.2 Estimates of sampling and stochastic uncertainty 54

8.3 Estimates of calibration uncertainty 60

8.4 Estimates of five sources of uncertainty 66

8.4.1 Results from reform 1, using 2 seed sets 67

8.4.2 Results from reforms 1-6 69

8.4.3 Results from reforms 7-11 74

8.4.4 General comparisons of all 11 reforms 77

9 Conclusions 81

10 References 85

A Appendices 89

A.1 Figures 89

A.2 ANOVA tables, expected mean squares, and variance estimates 93

A.3 Tables of results from individual experiments 100

Table of Contents

A.3.1 Experiments using fewer than 10 MATH-CPS model modules 100

A.3.2 Experiments with sampling and stochastic uncertainty 102

A.3.3 Experiments with sampling, stochastic and calibration uncertainty 106

A.3.4 Experiments with five sources of uncertainty 109

A.4 Overview of relevant MATH-CPS modules 121

A.5 Comparison of methods for determining FSP participation probabilities 127

Table of Contents

List of Figures

1 Two-cell calibration weighting factors for AFDC-Basic. 89

2 Two-cell calibration weighting factors for AFDC-UP. 90

3 Comparison of calibration effect, when calibration is done once using MPR seed, to calibration of each run. 91

4 Comparison of calibration effect, when calibration is done once using weighting factors based on the median of 80 runs, to calibration of each run. 92

List of Tables

1 Summary of Measurements of Five Sources of Uncertainty, Using Sample Reforms 1-6 xxiv

2 Summary of Measurements of Five Sources of Uncertainty, Using Sample Reforms 7-11 xxv

3 MATH-CPS model modules 22

4 Some baselaw target quantities of interest in the MATH-CPS model, and the extent to which they are achieved. 31

5 Public Assistance Eligibility and Participation Summary . . . 35

6 Output table showing entries for which variances were calculated 40

7 Estimators and variances for calibration, unemployment, and income allocation effects and their 2-way and 3-way interactions 48

8 Summary results for experiments varying the number of MATH-CPS model modules run 52

9 Summary results from experiments using stochastic and sampling variability, from a nested design. 56

10 Summary results from experiments using stochastic and sampling variability, from a crossed design 57

11 Summary results from experiments using stochastic and simulation variability, from a combination of a nested and a crossed design 58

12 Details of 2 cell calibration method, based on one seed and the full sample 61

13 Summary results from calibration experiments. 62

14 Details of 3 cell calibration method, based on one seed and the full sample 63

15 Comparison of baselaw targets and microsimulation estimates using two different calibration methods 64

16 Details of 3 cell calibration method when calibration is done separately for each run 65

17 Comparison of weighting factors in 3 cell calibrations based on the MPR seed versus the median of 40 runs 66

18 Summary results from the full 5-factor model under reform 1, using two different seed sets. 68

19 Description of reforms under which the full MATH-CPS model was run 70

20 Estimates of main, 2-way and 3-way effects in MATH-CPS model reforms 1-6. 71

21 Estimates of simulation and sampling variability in MATH-CPS model reforms 1-6 71

22 Percentage of posterior variance in a full run due to major sources of error in MATH-CPS model reforms 1-6 72

23 Estimates of main, 2-way and 3-way effects in MATH-CPS model reforms 7-11. 75

24 Estimates of simulation and sampling variability in MATH-CPS model reforms 7-11 75

25 Percentage of posterior variance in a full run due to major sources of error in MATH-CPS model reforms 7-11 76

Table of Contents

26 Estimates of the standard error and coefficient of variation in all 11 MATH model reforms. Numbers are for a single run, with full run numbers in parentheses. 78

27 Percentage of posterior variance in a full run due to major sources of error in all 11 MATH model reforms. 78

28 Sums of squares from ANOVA using a nested design, for experiments with sampling, and stochastic simulation uncertainty. 93

29 Expected mean squares from ANOVA using a nested design, for experiments with sampling and stochastic simulation uncertainty. 93

30 Variance components estimators from a nested design, for experiments with sampling and stochastic uncertainty. 93

31 Sums of squares from ANOVA using a crossed design, for experiments with sampling, and stochastic simulation uncertainty. 94

32 Expected mean squares from ANOVA using a crossed design, for experiments with sampling and stochastic simulation uncertainty. 94

33 Variance components estimators from a crossed design, for experiments with sampling and stochastic uncertainty. 94

34 Sums of squares from ANOVA for experiments with sampling, stochastic simulation, and calibration uncertainty 95

35 Expected mean squares from ANOVA for experiments with sampling, stochastic simulation, and calibration uncertainty . 95

36 Variance components estimators for experiments with sampling, stochastic simulation, and calibration uncertainty 96

37 Sums of squares from ANOVA for experiments with sampling, stochastic simulation, calibration, unemployment, and income allocation uncertainty 97

38 Expected mean squares from ANOVA for experiments with sampling, stochastic simulation, calibration, unemployment, and income allocation uncertainty 98

Table of Contents

39 Variance components estimators for experiments with sampling, stochastic simulation, calibration, unemployment and income allocation uncertainty. 99

40 Results from running just the FSTAMP reform part of the MATH-CPS model incorporating sampling and stochastic variability 100

41 Results from running the FSTAMP baselaw and reform parts of the MATH-CPS model incorporating sampling and stochastic variability 100

42 Results from running 8 modules of the MATH-CPS model incorporating sampling and stochastic variability 101

43 Results from running 9 modules of the MATH-CPS model incorporating sampling and stochastic variability 101

44 Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds nested, and seed set A 102

45 Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds nested, and seed set B 102

46 Results from running 10 modules of the MATH-CPS model, using a jackknife of 19 random groups, 4 seeds nested, and seed set A 103

47 Results from running 10 modules of the MATH-CPS model, using a jackknife of 19 random groups, 4 seeds nested, and seed set B 103

48 Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds crossed, and seed set A 103

49 Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds crossed, and seed set B 104

50 Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds nested, and seed set A 104

Table of Contents

51 Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds nested, and seed set B . . . 104

52 Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds crossed, and seed set A . . . 105

53 Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds crossed, and seed set B . . . 105

54 MATH-CPS model results, calibrating AFDC totals based on weighting factors from the MPR seed, separating households into 2 groups. A jackknife of 10 random groups was used with 4 seeds nested . . . 106

55 MATH-CPS model results, calibrating AFDC totals based on weighting factors from the MPR seed, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested . . . 107

56 MATH-CPS model results, calibrating AFDC totals on each run, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested . . . 107

57 MATH-CPS model results, calibrating AFDC totals based on median weighting factors, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested 108

58 MATH-CPS model results, calibrating AFDC totals on each run, separating households into 3 groups. A jackknife of 8 rotation groups was used with 4 seeds nested . . . 108

59 MATH-CPS model results under reform 1 using 5 sources of uncertainty, and seed set A . . . 109

60 MATH-CPS model results under reform 1 using 5 sources of uncertainty, and seed set B . . . 110

61 MATH-CPS model results under reform 2 using 5 sources of uncertainty, and seed set A . . . 111

62 MATH-CPS model results under reform 3 using 5 sources of uncertainty, and seed set A . . . 112

Table of Contents

63 MATH-CPS model results under reform 4 using 5 sources of uncertainty, and seed set A 113

64 MATH-CPS model results under reform 5 using 5 sources of uncertainty, and seed set A 114

65 MATH-CPS model results under reform 6 using 5 sources of uncertainty, and seed set A 115

66 MATH-CPS model results under reform 7 using 5 sources of uncertainty, and seed set A 116

67 MATH-CPS model results under reform 8 using 5 sources of uncertainty, and seed set A 117

68 MATH-CPS model results under reform 9 using 5 sources of uncertainty, and seed set A 118

69 MATH-CPS model results under reform 10 using 5 sources of uncertainty, and seed set A 119

70 MATH-CPS model results under reform 11 using 5 sources of uncertainty, and seed set A 120

Table of Contents

Executive Summary

A. Introduction

In a 1991 review, a National Academy of Sciences panel on the uses of microsimulation modeling for informing policy decisions identified the "lack of regular and systematic model validation" as one of the two major deficiencies that must be addressed. As part of its recommendations on model validation, the panel urged investigation of the sources of uncertainty in model estimates. Toward this end, FCS funded error analysis research on its QC- and SIPP-based FSP models (see Zaslavsky and Thurston, 1995). FCS also funded research on its CPS-based FSP model through Task 29 of the current FSP microsimulation contract. The product of Task 29 is the report entitled "Error Analysis of MATH-CPS Model Estimates" by Thurston and Zaslavsky of Datametrics Research, Inc.

In this executive summary, we present an overview of the Datametrics report. Our purpose is to summarize the report in a way that makes the results accessible to the average user of the FSP microsimulation models at FCS and MPR.

B. The Purpose of Error Analysis

When using microsimulation model estimates, we are always uncertain how close a model estimate is to the true effect of a reform proposal. Sometimes we are more uncertain than others. For example, when our model estimate is based on a small sample of the FSP population, we are more uncertain than when our estimate is based on a large sample. Sometimes a particular source of uncertainty contributes more to total error than it does at other times. For example, uncertainty caused by the use of imputed asset holdings is more important for reforms involving asset limits than for other types of reforms.

The purpose of the error analysis performed by Thurston and Zaslavsky was to identify sources of uncertainty and estimate their magnitude. The error analysis is essentially a sensitivity analysis. The analysis seeks to measure how sensitive model estimates are to various influences that cause uncertainty in the estimates. Sensitivity analysis can help us understand how different a model estimate might have been if:

- the Census Bureau had drawn a different sample
- we had used a different set of pseudo-random numbers to model the participation decision
- we had used an alternative specification for the asset imputation equation
- the allocation of calendar year income to months during the year had been modeled differently
- we had calibrated the baseline simulation using a different model of participation
- we had "aged" the database using a different prediction for the rate of wage erosion for adults with low levels of education

Answers to questions such as these are valuable because they inform both the use and the development of the model.

Knowing something about the uncertainties associated with a model estimate can help a user decide about how best to refine the estimate or add other information to it. For example, if a user knows that an estimate has large uncertainty due to sampling variability, he or she may decide to run the reform on two different samples (e.g., using both the January 1992 and the January 1994 SIPP data), and then use an average of the two resulting estimates. Alternatively, if the user knows that an estimate is very sensitive to the unemployment rate reflected in the model database, he or she may decide to run the model using two or three different unemployment rate assumptions, and then use an average of these estimates, with the average based on OMB projections of likely unemployment rates in the time period of interest.

Knowing something about the uncertainties associated with a model esti-

Knowledge of the uncertainties associated with model estimates is also important for guiding model development activities. Model development efforts can be focused on model components that bring the most improvement in the accuracy and reliability of the estimates. For example, if a key class of estimates (e.g., those for vehicle reforms) are known to be sensitive to the method of allocating calendar year income to months of the year, then we may decide to focus some model development effort on reducing our uncertainty about the best way to perform this allocation.

C. The Scope of the Datametrics Study

The report by Thurston and Zaslavsky describes both the development of an error analysis methodology, as well as the application of the methodology to a set of eleven sample reforms. Most of the work was devoted to development of the technique; the work was not devoted to any sort of systematic assessment of the MATH-CPS model. The application work only begins to scratch the surface of the analysis that can be done using the techniques developed. The analysis of the eleven sample reforms does provide some interesting and valuable knowledge about uncertainties in the estimates for these reforms. However, the results cannot be generalized to other reforms. Also, note that the sources of uncertainty and the sample reforms analyzed were both selected to illustrate different types of uncertainty, rather than to comprehensively address all issues of uncertainty.

D. Sources of Uncertainty in MATH-CPS Model Estimates

The report classifies sources of uncertainties into four broad categories: sampling, stochastic simulation, model specification, and model parameters.

Sampling Uncertainty. Sampling uncertainty is caused by the use of data from a representative sample, rather than from the entire population of interest. If another sample, drawn by the same method, is used for the model database, the estimates will be different.

Stochastic Simulation Uncertainty. Stochastic simulation uncertainty is caused by the use of pseudo-random numbers to simulate household decisions or to add random variability to an imputed variable. For example, random numbers are used to simulate whether or not a household that experiences a decline in FSP benefit decides to stop participating. Another example is the use of random numbers for the imputation of the month of the year a person is simulated to begin employment (if the person is employed for less than 52 weeks). For any use of a random number, if a different set of random numbers is used, the model estimates will be different.

Model Specification Uncertainty. Model specification uncertainty is caused by our uncertainty about how best to model the FSP using the available data. We are never certain about which model specifications most closely reflect reality. For example, there may be two equally plausible ways to model the composition of FSP filing units. We are uncertain which method is best, but we choose one. If we had chosen the other, the resulting model estimates would be different.

Other examples of model specifications that produce uncertainty include:

- The method used to correct for under reporting of various income sources. Many different approaches are possible and we do not know which is the most realistic one.
- The regression models used to impute assets and expenses. Imputed variables are typically modeled as a function of a set of predictor variables. Alternative models that specify the imputed variable as a function of a different set of predictor variables may also be plausible.
- The method used to "age" the database to a future time period.
- The method used to calibrate the baseline participation model so that the characteristics of the simulated population of FSP recipients are similar to those of the actual population.
- The method used to calibrate the asset and expenses imputations so that the average assets and expenses of the simulated population of FSP units are similar to the averages observed in administrative data.
- The modeling of FSP eligibility rules. Not every provision of the FSP regulations is captured in the model specification. Some provisions are modeled in great detail, while others are captured only roughly. We are always uncertain about which eligibility rules to model and how best to model them with the available data. For many FSP provisions, several different modeling approaches are reasonable.

Two sources of model specification uncertainty are analyzed in the report. The first model specification analyzed is the specification for calibrating the total number of AFDC participants in the baseline simulation. In the development of the current MATH-CPS baseline, the total number of AFDC participants and benefits projected for April 1996 could not be reached. MPR decided to simply select all eligible units to participate. The resulting baseline seemed reasonable, even though it did not meet the desired targets for total participants and benefits. Other methods of calibrating the baseline total AFDC participants and benefits are possible. For the error analysis, Thurston and Zaslavsky contrasted the official algorithm with an alternative specification for calibrating the baseline total AFDC participants and benefits. They used a re-weighting method that increased the weights of households eligible for AFDC and decreased the weights of households ineligible for AFDC. Weights were adjusted differentially based on AFDC benefit, in order to ensure that targets for both participants and benefits were met.

The second model specification analyzed in the report is for the method of allocating annual earnings across months. The current model allocates earnings evenly across all months worked. This model may be realistic for many persons, but for others, work may have been distributed unevenly across time. For example, some people may work overtime on a seasonal basis, say during the summer. Other people might work full time during the first half of the year, but then change jobs and work only part time during the remainder of the year. If we model earnings as unevenly allocated across the months of the year, some of the FSP-eligible households in the current baseline during the simulation month (April) would not be eligible in the alternative baseline, and some who are not eligible in the current baseline would be eligible in the alternative baseline. Thus, with different households in the baseline, reform estimates would be different.

To explore the sensitivity of model estimates to the model specification for the timing of earnings during the year, Thurston and Zaslavsky used an alternative model specification in which earnings are modeled to fluctuate from month to month during the year. During the simulation month, earnings were modeled to be between 0.5 and 1.5 times the average earnings over all months. A uniform random number was used to determine the ratio of earnings in the simulation month to the earnings in the average month. This alternative earnings allocation model was only used for households with total income below the poverty level.

Model Parameter Uncertainty. Many of the model specifications in MATH-CPS use model parameters. Model parameter uncertainty is caused by our uncertainty about the correct value of the parameters. For example, we have devised a method for aging the income on the database to represent income received in a future year. The method calls for a parameter that specifies the expected inflation rate for the U.S. dollar. We are uncertain about what the actual inflation rate will be, but we choose one estimate (the one produced by OMB) for use in "aging" income amounts. If we use a different estimate for the inflation rate (e.g., one produced by a Wall Street firm), the estimates produced by the model will be different.

Other examples of model parameters that produce uncertainty include:

- Calibration targets (e.g., percent of FSP units with a disabled person).
- Demographic and economic aging factors.
- Predictions for future AFDC maximum benefit levels.
- Coefficients of the regression equations used to impute assets and expenses. Because the regression coefficients were estimated using a population sample, they are subject to sampling error and other types of error.

One source of model parameter uncertainty is analyzed in the report: the expected unemployment rate for the simulation month. The unemployment rate in the future simulation month is an unknown parameter. A model for changing the unemployment rate from the one reflected in the data was developed by FCS and MPR, but for the April 1996 MATH-CPS database, we chose not to use the model but instead retain the unemployment rate reflected in the CPS data for 1992. In the report, Thurston and Zaslavsky classify uncertainty about the future unemployment rate as a "macro effect." A macro effect is a broad change in the state of the world, such as in the unemployment rate or in overall wage levels.

For the error analysis in the report, Thurston and Zaslavsky ran reform simulations using an alternative unemployment rate. The alternative unemployment rate was achieved by changing the allocation of weeks unemployed during the year to particular months. In the current procedure, the beginning month of unemployment is randomly chosen. In the alternative procedure, for some persons, the unemployment spell is forced to not occur during the simulation month. Thus, some persons who are normally simulated to be unemployed during the simulation month are simulated in the alternative model as being employed during the simulation month. The alternative specification implicitly models changes in the unemployment rate as mainly due to changes in the unemployment spell length for people who are unemployed at some point during the calendar year. The alternative procedure decreased the unemployment rate in the simulation month from 6.4 percent to 4.3 percent.

Table of Contents

Note that the method used to change the unemployment rate also involves model specification uncertainty, since many other methods of changing the unemployment rate are also plausible. However, the purpose of altering the unemployment rate was to explore parameter uncertainty (in this case, due to a macro effect).

E. Methodology

The main approach used by Thurston and Zaslavsky to estimate the magnitude of uncertainties in the model estimates was a sensitivity analysis using a Bayesian framework. A sample reform simulation was run under different conditions, thus creating variability in the resulting model estimates. Sampling variability was created using different subsamples of the database (i.e., jackknife resampling). Stochastic variability was created using different sets of seeds for the random number generator. Model specification variability and model parameter variability were created by running the model using both the current (official) algorithm and one alternative algorithm for each of the model uncertainty factors.

Depending on the particular way that jackknife replicates and random number seeds are combined, there are several possible ways to run the simulations (i.e., several possible experimental designs). Thurston and Zaslavsky made use of two different designs, one nested and one crossed.

For either experimental design, the contributions of each altered condition to the variance of the model estimate can be summarized using analysis of variance (ANOVA). Mean square errors (from the ANOVA table) were calculated using outputs from the different simulation runs. These calculated mean square errors were then used to estimate the parameters of a theoretical model for the effect of each type of uncertainty on the model estimate.20

The theoretical model expresses the model output of interest (the percent change in FSP participants or benefits) as a linear function of the "influences" of each uncertainty source. The theoretical model was used in two ways. First, expressions for expected mean square errors were derived from the theoretical model and then equated to the mean square errors from the ANOVA table for the actual MATH model estimates. Solving the equations for the variance components produced an estimator for each variance component. Second, the theoretical variance components model was used in a Bayesian framework to derive expressions for the posterior variances of the quantities of interest (primarily the posterior variance of the percent change in benefits). The posterior variance of the percent change in benefits is ex-

Table of Contents

pressed as a sum of variance components, each of which represents either one of the five modeled sources of uncertainty, or the interaction of two or more of the sources.

In a Bayesian framework, all uncertainties are described by probability statements. For example, if we are not sure which of two specifications for a particular part of the model is correct, before seeing the data we might say that each of the two has a 50 percent probability of being the correct specification. In this framework, the posterior variance (where "posterior" means "after seeing and analyzing the data") summarizes our uncertainty about an estimand after using all available information and taking into account all remaining uncertainties. In practice, when analyzing something as complex as a microsimulation model we do not take into account every conceivable source of uncertainty but restrict our attention to those we consider important and amenable to analysis.

The use of a Bayesian framework was important because it allowed the consideration of both non-random effects (model specifications and model parameters) and random effects (sampling and stochastic simulation) as contributors to posterior variance. This allowed estimation of the relative importance of each factor. Without a Bayesian approach, point estimates of outputs can be compared under different model specifications. This is useful for examining the impact of a particular feature of the model specification. However, we would be unable to ascertain the degree to which model specifications and parameters contribute to the total uncertainty of the estimate compared to the degree to which purely random elements (such as sampling) contribute.

F. Summary of Key Results

In this section, we summarize some of the key results presented in the report. We divide these into three groups: (1) results from experiments measuring only stochastic and sampling uncertainty using one sample reform; (2) results from experiments measuring all five sources of variability using eleven different sample reforms; and (3) other key results.

1. Measurements of Sampling and Stochastic Simulation Uncertainty Using One Sample Reform

Experiments showed that sampling uncertainty was somewhat difficult to measure reliably. We had expected variability caused by sampling to be easy to measure well because it is easy to measure by itself. In the absence of other sources of uncertainty, sampling uncertainty is easy to measure reasonably well by simply drawing subsamples of the CPS sample (e.g., using jackknife resampling). The difficulty in measuring sampling uncertainty arose in measuring both sampling and stochastic simulation uncertainty simultaneously.

Estimates for sampling and stochastic variance for a sample reform simulation were sensitive to several experimental conditions: the experiment design (crossed or nested), the set of random number seeds used, the number of jackknife replicates used, and the composition of the jackknife replicates (rotation groups or random groups). For sample reform 1 (see Tables 1 and 2 for description of the reform), estimates of the percent change in benefits were about 9 percent. Estimates of sampling standard error varied substantially: 8 estimates were zero, 14 estimates were in the 0.1 to 0.4 range, and 6 estimates were in the 0.4 to 0.7 range. Estimates of the stochastic error varied from 0.42 to 1.03, with most in the 0.5 to 0.8 range. Estimates of the percent variance due to stochastic error ranged from 38 percent to 100 percent, with most in the 70 to 100 percent range.

While the sampling and stochastic error estimates do seem sensitive to the experimental conditions, they are all small relative to the impact estimate of about 9 percent increase in total FSP benefits. Viewed relative to the impact estimates, the estimates for sampling and stochastic error do not vary wildly. Estimates for sampling error range from 0 to at most 8 percent of the estimate. Estimates for stochastic simulation error range from 0 to at most 11 percent of the estimate. All of the estimates are reliable in that they all tell us that both sampling and stochastic error are quite small for sample reform 1. Estimates for other reforms might prove less sensitive to the experimental conditions.

Although sampling and stochastic simulation error estimates varied considerably, some of the estimates were judged to be better than others. For example, we believe that the nested design produced better estimates for stochastic simulation uncertainty, while the crossed design produced better estimates for sampling uncertainty.

Also, although sampling and stochastic simulation error proved somewhat difficult to measure, much of the difficulty appeared to be in the allocation of error between these two components, which is less crucial than estimation of the total error produced by both components.

2. Measurements of Five Sources of Uncertainty Using Eleven Sample Reforms

Experiments using the full five-factor error model, with eleven different sample reforms provide some of the most interesting results in the report. These results are summarized in Tables 1 and 2. Table 1 contains results from the first six reform simulations, all of which produced estimates with little variability. For these estimates, the standard error ranged from 7 to 15 percent of the estimate. Table 2 contains results from five subsequent reform simulations, all of which produced estimates with substantial variability. For these estimates, the standard error ranged from 63 to 193 percent of the estimate.

The top section of Tables 1 and 2 contains estimates for the percent change in FSP benefits. The overall estimate and the estimates for each non-random effect are listed for each reform. (Random effects are those due to sampling and stochastic simulation.) The estimates for the non-random effects show the effect of using the alternative calibration, unemployment rate, or income allocation model, compared to using the versions in the official model. For example, for reform 1, the estimate for the percent change in FSP benefits is 9.29 percent (8.47 + 0.82) using the official (MPR) calibration, and 7.65 percent (8.47 - 0.82) using the alternative (Datametrics) calibration. (The estimates for the non-random effects were calculated using a flat prior; i.e., the official specification and the alternative specification were assumed to be equally plausible).

The middle two sections of Tables 1 and 2 contain the most striking results in the report. Estimates of the standard error of the overall estimate, by error source are presented in both absolute terms, and as a percentage of the overall estimate. The error estimates show that the each of the five measured sources of error can be substantial in magnitude, for at least some of the sample reform simulations. By "substantial" in magnitude, we mean that the error is large relative to the estimate itself. Sampling error was important for reforms 8, 9, and 10. Stochastic error was important for reforms 7, 8, 10, and 11. Error due to uncertainty about the best AFDC calibration procedure was important for reforms 8, 9, 10, and 11. Error due to uncertainty about the correct unemployment rate was important for reforms 7, 8, 10, and 11. Error due to uncertainty about the best model for allocating calendar year income to months was important for reforms 7, 10, and 11.

In some cases, a single source of error by itself produces a wide confidence interval. For example, if sampling were the only source of error, our estimate of 0.17 percent change in FSP benefits for reform 10 would have a 70 percent confidence interval running from about 0 to about 0.35 percent. Likewise, if error due to uncertainty about the correct unemployment rate were the only source of error, our estimate of 0.29 percent for reform 11 would have a 70 percent confidence interval running from about -0.05 to about 0.63 percent.

While each of the five sources of error proved to be important for some of the reform simulations, each of the sources of error also proved to be unimportant for other reform simulations. For reforms 1 through 6, none of the sources of error produced a standard error greater than 15 percent of the estimate.

In Tables 1 and 2, we also present the error component estimates as percentage of the total error measured. These estimates allow us to compare the importance of each source of error (relative to other sources of error) across reforms that have very different levels of total error. These estimates nicely illustrate what model users know intuitively: the sensitivity of the estimate to different influences depends entirely on the reform.

Sampling uncertainty was important (relative to other sources) for reforms 1, 5, 8, 9, and 10, but not at all important for other reforms. Stochastic simulation uncertainty was important for reforms 3, 4, 6, 7, and 11. Calibration uncertainty was important for reforms 1 and 5, and moderately important for reforms 2, 3, 4, and 8. Uncertainty about the unemployment rate was very important for reforms 2 and 6 (increases to the earnings deduction), and moderately important for reforms 3,10, and 11. The model for allocating calendar year income across time was generally not important relative to other sources of uncertainty.

The specification for calibrating AFDC participation totals was important (relative to other sources of error) for several reforms even when AFDC participation does not seem directly related to these reforms. For example, reform 4 excludes the first vehicle for all households; using the official calibration for AFDC produced an estimate of 5.67 percent change in FSP benefits, while using the alternative calibration produced a 4.96 percent change in benefits. Other MATH model calibration procedures (e.g., for the vehicular assets imputations, and for baseline FSP participation) might prove to be even more important sources of uncertainty in estimates for vehicle reforms (as well as for other types of reforms).

3. Other Key Results

The report contains many additional valuable results; we describe only a few of them here.

The use of multiple random number set can reduce error substantially for some reforms. Currently, MATH-CPS model estimates are produced using only one set of random numbers. The one set used may or may not produce estimates that are typical of those that would be produced using other random number sets. By running the model using multiple random number streams and averaging the resulting impact estimates, we can reduce stochastic simulation uncertainty. For the sample reforms run for this report, the last line of Tables 1 and 2 shows the percent reduction in the standard error of the reform estimate due to using 14 different random number sets for each household. Although stochastic simulation uncertainty can be virtually eliminated by using just a few different random number sets, the overall reduction in uncertainty that can be achieved depends on the degree to which stochastic simulation uncertainty contributes to total posterior variance and the magnitude of total posterior variance. For reforms 7 and 11, the use of multiple random number sets would substantially decrease total uncertainty.

The particular AFDC calibration method is important. In developing the alternative method used to calibrate total AFDC participants and benefits, Thurston and Zaslavsky were uncertain how best to do it, even after having decided on a reweighting scheme over other methods. They tried multiple variations of the reweighting specification, and the results from these variations nicely illustrate the idea of model specification uncertainty. The alternative estimates for reform 1 ranged from 0.66 to 1.8 percentage points below the 9.3 percent obtained using the official MPR calibration. Estimates were sensitive to several factors, including the particular random number set used when tabulating AFDC totals to develop the weighting factors. Obtaining the best results required the use of reweighting factors developed from the median number of AFDC eligibles.

The AFDC calibration experiments showed that when we try to obtain more complete consistency between model totals and control totals, we obtain substantially different results for some reforms. These results suggest that more research about the causes of discrepancies between survey and administrative data might prove worthwhile. The error analysis results thus confirm the puzzlement that many analysts have had about this issue.

Measures of sampling error can inadvertently capture stochastic simulation error. If stochastic simulation error is not eliminated by running

Table of Contents

the model using multiple random number seeds, then measures of sampling error also measure stochastic error. Experiments that varied the number of MATH-CPS modules in the run sequence demonstrated this result. An example best explains the phenomenon. Suppose that we used a very simple imputation for shelter expenses in which shelter expenses are set equal to 0.3 times income, plus a normally distributed random number with zero mean:

$$\text{shelter expense} = 0.3 \times \text{income} + Z$$

The random number, Z, is used to simulate variation in expenses that cannot be captured by our simple 0.3-times-income model. This addition of a random variation term is done for most equation-based imputations in the MATH-CPS model.

Using our imputed shelter expenses, we can estimate average shelter expense by calculating the mean shelter expense value on our database. Our estimate of average shelter expenses is subject to sampling error; other samples with different average income would give us different estimates of the average

Table of Contents

G. Conclusions

The work presented in the report is a valuable contribution to understanding the sources of uncertainty in MATH-CPS model estimates. The techniques developed for this work are widely applicable to many of the modeling activities supported by FCS. Much of the methodology developed is generalizable to the SIPP- and QC-based FSP models. The knowledge gained through this error analysis will help to inform any error analysis work needed by FCS in the future. For example, this work will likely inform the measurement of uncertainty that will be important in using CPS and SIPP data for producing state-level impact estimates.

Table of Contents

Table 1: Summary of Measurements of Five Sources of Uncertainty, Using Sample Reforms 1-6

	Sample Reform Number					
	1	2	3	4	5	6
Estimates for percent change in FSP benefits						
overall	8.47	10.44	2.84	5.31	26.06	3.08
calibration procedure	-0.82	-0.50	0.09	-0.36	-1.71	0.07
unemployment rate	-0.08	0.80	-0.05	-0.05	-0.12	0.28
income allocation model	0.04	-0.43	-0.04	-0.02	0.02	-0.04
interactions	0.02	-0.03	0.00	0.01	0.02	0.00
Standard error of overall estimate, by component						
total	1.26	1.11	0.19	0.75	2.63	0.39
sampling	0.67	0.00	0.04	0.00	1.77	0.00
stochastic	0.53	0.35	0.12	0.59	0.66	0.23
calibration procedure	0.83	0.51	0.09	0.37	1.72	0.07
unemployment rate	0.37	0.81	0.09	0.24	0.57	0.29
income allocation model	0.15	0.44	0.05	0.14	0.26	0.07
interactions	0.05	0.07	0.02	0.03	0.06	0.04
Standard error of overall estimate, as a percent of estimate						
total	15%	11%	7%	14%	10%	12%
sampling	8%	0%	2%	0%	7%	0%
stochastic	6%	3%	4%	11%	3%	8%
calibration procedure	10%	5%	3%	7%	7%	2%
unemployment rate	4%	8%	3%	4%	2%	9%
income allocation model	2%	4%	2%	3%	1%	2%
interactions	1%	1%	1%	1%	0%	1%
Variance components, as a percent of total measured variance						
total	100%	100%	100%	100%	100%	100%
sampling	29%	0%	5%	0%	45%	0%
stochastic	18%	10%	41%	62%	6%	36%
calibration procedure	43%	21%	22%	24%	43%	4%
unemployment rate	9%	53%	23%	10%	5%	56%
income allocation model	1%	16%	7%	3%	1%	3%
interactions	0%	0%	1%	0%	0%	1%
Reduction in standard error of overall estimate when using multiple seeds						
	9%	5%	25%	36%	3%	22%

Reform #1: Increase asset limits to $5,000; reduce shelter deduction cap by 50 percent; remove cap on dependent care deduction

Reform #2: Increase earnings deduction to 50 percent

Reform #3: Remove cap on shelter deduction

Reform #4: Exclude entire value of first vehicle

Reform #5: Exclude all assets

Reform #6: Increase earnings deduction to 50 percent for AFDC households

Table of Contents

Table 2: Summary of Measurements of Five Sources of Uncertainty, Using Sample Reforms 7-11

	Sample Reform Number				
	7	8	9	10	11
Estimates for percent change in FSP benefits					
overall	0.31	0.22	0.30	0.17	0.29
calibration procedure	-0.03	0.08	0.06	-0.06	-0.05
unemployment rate	-0.02	-0.06	0.00	-0.15	-0.08
income allocation model	-0.00	0.01	-0.00	-0.03	0.00
interactions	0.01	0.00	-0.00	0.02	0.01
Standard error of overall estimate, by component					
total	0.32	0.17	0.19	0.32	0.56
sampling	0.00	0.13	0.16	0.18	0.00
stochastic	0.30	0.05	0.04	0.16	0.42
calibration procedure	0.06	0.08	0.07	0.06	0.10
unemployment rate	0.08	0.06	0.04	0.16	0.34
income allocation model	0.07	0.02	0.02	0.09	0.11
interactions	0.02	0.01	0.02	0.02	0.03
Standard error of overall estimate, as a percent of estimate					
total	105%	80%	63%	190%	193%
sampling	0%	60%	54%	109%	0%
stochastic	97%	21%	13%	96%	145%
calibration procedure	18%	36%	24%	39%	34%
unemployment rate	27%	30%	13%	99%	117%
income allocation model	22%	9%	6%	57%	39%
interactions	7%	5%	7%	15%	11%
Variance components, as a percent of total measured variance					
total	100%	100%	100%	100%	100%
sampling	0%	57%	74%	33%	0%
stochastic	86%	7%	5%	25%	56%
calibration procedure	3%	21%	14%	4%	3%
unemployment rate	7%	14%	4%	27%	36%
income allocation model	4%	1%	1%	9%	4%
interactions	0%	0%	1%	1%	0%
Reduction in standard error of overall estimate when using multiple seeds					
	65%	4%	8%	16%	33%

Reform #7: Asset limit = $2,150 for both elderly and non-elderly units

Reform #8: Shelter deduction = shelter expenses over 35% of gross income

Reform #9: Earnings deduction = 75% for units participating in AFDC-UP

Reform #10: Earnings deduction = 34% for units on FSP for 6 months or less; otherwise, no earnings deduciton

Reform #11: Exclude $10,000 of first vehicle; reduce asset limit to $1,075 for non-elderly units on FSP for more than 6 months

Table of Contents

1 Introduction

Microsimulation models are one of the main tools used to estimate the cost and other impacts of proposed changes in Federal programs. In these models, behavior is simulated at the level of the individual unit, hence the name *microsimulation*. By taking into account the specific characteristics of each unit, microsimulation models have the potential to better estimate effects of small changes in policy and program provisions, and their differential impact among population subgroups, than can models that operate at a more aggregate level (Citro and Hanushek, 1991). In addition, microsimulation models can address questions such as who gains and who loses from a proposed policy change.

Typically microsimulation models as currently implemented give point estimates with no estimates of variability. However there are many sources of uncertainty in these models including: sampling, stochastic simulation, choice of model parameters, model specification including calibration, and macro effects (which can potentially affect any household) such as unemployment.

A panel of experts convened by the Committee on National Statistics at the National Research Council recommended in 1991 that validation studies of microsimulation models be conducted by independent contractors. One type of validation study recommended was a sensitivity analysis, in which outputs from running a microsimulation model under changed sets of conditions are compared (Citro and Hanushek, 1991).

We use sensitivity analysis to model some of the uncertainties in the Micro Analysis of Transfers to Households (MATH-CPS) model. The MATH-CPS model is one of three microsimulation models currently used to estimate the impacts of changes to the Food Stamp Program (FSP). Outputs of particular interest from these models include comparisons of benefits received under the current law with those received under a changed set of program rules. We summarize uncertainties in these comparisons by estimating the relative contribution of each error source to posterior variance.

We have modeled five sources of uncertainty in the MATH-CPS model, as described in Section 6. A summary of the sources of uncertainty which we modeled appears at the beginning of Section 7.1.

Table of Contents

2 Sources of uncertainty in microsimulation models

Microsimulation models use data from a base dataset, which is usually obtained from survey data or administrative records. Often, not all variables needed for the model are included in the base dataset. The other variables needed may be estimated using information from a model or an auxiliary dataset. In some microsimulation models, the data is aged to a future time. Decisions are then simulated at the level of individual units under the current set of laws as well as under one or more sets of proposed policy changes. Finally, outputs of interest are aggregated within subgroups of interest.

There are many sources of uncertainty in microsimulation models. Sampling uncertainty refers to uncertainty due to the fact that the data is based on a sample of units. Sampling uncertainty arises from use of the base dataset and from the auxiliary dataset, if the latter is used. In addition, model parameters may be based on supplementary studies, which themselves are subject to sampling uncertainty.

Stochastic simulation uncertainty arises when random numbers are used in the model. Random numbers are often used to simulate decisions of individual units in the dataset. Stochastic simulation uncertainty also arises when missing data are imputed, and may arise in statistical matching.

There are many underlying models implicitly or explicitly specified in microsimulation models. One example is the assumption that the database is free of deficiencies. However there may be discrepancies between the database and what is believed to be correct. For example, it may be known that some percentage of the population receives a particular form of public assistance, but the percentage in the database doesn't match this number. This leads to uncertainty about how to correct the database deficiencies. Another example is when a decision simulated for an individual unit in the database is modeled to be a function of a set of predictor variables. An alternate model could be plausible which specifies that the decision is a function of a different set of predictor variables, where the two sets may contain the same, overlapping, or totally different variables. We use the term model specification uncertainty to denote uncertainty about which model most closely reflects reality.

Model specification uncertainty arises in nearly all aspects of microsimulation modeling. During the database creation stage, statistical matching is often used for imputation of missing values. In statistical matching, a typical model assumes conditional independence: conditional on the variables common to both datasets, the variables in the base dataset are assumed to be independent of the variables in the auxiliary dataset. In an alternate model, the partial correlation between the variables which are just in the base dataset and the variables just in the auxiliary dataset, given the variables in both datasets, is assumed to have some non-zero value. This model would almost certainly give a different matched dataset.

The partial correlation coefficient cannot be determined from the dataset. Even if it was known, the resulting dataset from statistical matching is not identical to the data that would have been observed if there was no missing data. Whether or not the partial correlation coefficient is assumed to be 0, using just one set of imputed values doesn't reflect the uncertainty about the missing values. Use of multiple imputation, in which missing values are imputed several times resulting in multiple datasets, can allow for an estimate of the uncertainty due to the missing data (see Rubin, 1987, and Rubin, 1986).

Model specification uncertainty may arise in other stages of the creation of the base dataset. Aging of the dataset relies on a model, as does simulation of individual decisions. In some cases, application of program rules relies on a model.

Calibration is the process used to adjust the model in order to bring simulated totals or percentages closer to a projected total or percentage. If the data are calibrated, a model is specified for how the calibration is to occur. Calibration can be considered a type of model specification uncertainty. There are two aspects of calibration: how certain one is about the projected totals, and how to calibrate to these totals. Just as there can be many ways to specify a model, there can be many ways to specify how calibration can occur. Different methods of calibration may lead to different model outcomes.

The specification of a model usually involves model parameters. We refer to uncertainty about what the correct values of model parameters are as model parameter uncertainty. As an example, suppose an individual decision to participate in a public assistance program is modeled to be a function of a set of categorical predictor variables that specify an array of cells, used for table lookup. In the model implementation, an individual may be simulated to participate if a random number is less than the value in the cell relevant to that individual. There may be two plausible sets of values for the table and uncertainty about which set is closest to reality. Since many model specifications use model parameters, model parameter uncertainty may arise at almost any stage of the microsimulation model.

Some possible changes in the state of the world, such as the change in the unemployment rate, changes in overall wage levels, etc., may be reflected in changes in values of variables for many individuals in the file. We refer to uncertainty about these changes as uncertainty due to macro effects. Macro effects may potentially affect any individual in the file.

Finally, there is uncertainty involved in the choice of an outcome measure. Microsimulation models give many outputs, often including several different tables and many entries in each table. It may be of interest to have one overall measure of the effect of the reform, which would be based on some combination of cell outputs. These summary measures are called welfare measures. The outputs can be combined in many different ways, with uncertainty reflecting lack of consensus about how best to combine the outputs. Different ways of combining outputs may be preferable depending on what the purpose of the summary is. In theory, one measure may lead to the conclusion that the plan is beneficial as a whole, while another measure could lead to the opposite conclusion. Uncertainty about welfare measures could be estimated by the variability between different ways of summarizing the effects of the reform (i.e. between different ways of combining cell outputs).

Table of Contents

3 Approaches and results of error analysis in other studies

Citro and Hanushek (1991, p. 232) distinguish three approaches for validation of microsimulation models: external validation, sensitivity analysis, and variance estimation. In external validation, model outputs which typically simulate the effects of a changed policy, are compared with what actually happened. Estimates of what actually happened are usually based on administrative records, which themselves are subject to sampling variability. There are several difficulties in carrying out an external validation, including the possibility that the reform which was actually enacted was different from the reform used in the microsimulation model, and that changes in macroeconomic conditions in the intervening time period may have been different from what was modeled.

In a sensitivity analysis, different versions of model modules or different choices of model structures are used, and model outputs are compared across the different versions. The interest is in how sensitive the model outputs are to the changed set of conditions. Sensitivity analysis by itself cannot give information about which of the alternate versions are better, because it doesn't give a true reference point. However, sensitivity analysis can help guide future work by pointing out which model aspects need to be improved.

Citro and Hanushek (1991, p. 239) use the term "variance estimation" for estimation of sampling variability. They note that there is sampling variability both in the base data set and in any auxiliary datasets which may have been used for imputing missing variables, estimating model parameters, estimating control totals, estimating behavioral responses, etc.

Cohen (1991a) discusses how resampling methods can be used to estimate sampling variability in microsimulation models. He also mentions other potential sources of uncertainty in microsimulation models (p. 246-247), including choice of aging module, macroeconomic projections, and model parameters in regression. He suggests using a sensitivity analysis for different modules and different macroeconomic projections by selecting two or more versions of inputs, and combining this with bootstrap resampling (p. 250-251).

Of the relatively small number of studies which have addressed the uncertainty in or validity of microsimulation models, some focused on external validity, some focused on sensitivity analyses, and a few others addressed stochastic and/or sampling variability. Cohen (1991b) reviewed 13 stud-

Table of Contents

ies of microsimulation model evaluations. Of these, nine included external validation, eight included sensitivity analysis, and five gave suggestions for improvements to the models.

3.1 External validation studies of the MATH-CPS model

Two of the studies reviewed by Cohen (1991b) did an external validation of the MATH-CPS model: Doyle and Tripp (1989) and Beebout and Haworth (1989). Beebout and Haworth compared model outputs with actual changes which occurred after the 1977 Food Stamp Act was implemented. They concluded that the MATH-CPS model underestimated the impact of the Food Stamp Act by 1.8 to 12.8 percentage points (Beebout and Haworth, 1989, as cited in Cohen,1991b).

Doyle and Tripp (1989) evaluated the MATH-CPS model in two phases. In Phase I they evaluated model results using an unaged file, whereas in Phase II they evaluated the process of aging the database.

In Phase I, Doyle and Tripp compared outputs from the MATH-CPS model using the March 1985 CPS database, with August 1984 administrative data summaries, and with the MATH-SIPP model outputs using 1984 SIPP data. They note that the comparison database is also subject to sampling and other errors. Their results showed that many point estimates from the MATH-CPS model were close to administrative totals, including the total number of participants, total benefits paid, and the marginal distributions of participants along some dimensions such as gross monthly income and household size. However, they found that the distribution of simulated participants along other dimensions did not match the administrative data very well. In particular, they found that the model simulated food stamp participants among too many households with disabled non-elderly people, too many households with elderly people, too many households with earners, too few households with school-aged children, and too few non-elderly households on public assistance (Doyle and Tripp, 1989, Ch. 3). These problems remain with the MATH-CPS database with which we worked (Table 4). Doyle and Tripp also found that the MATH-SIPP model did not match administrative data better than the MATH-CPS model (p. 31).

Doyle and Tripp investigated a number of reasons for these discrepancies (Ch. 4). One problem they found was that there were not enough food stamp eligible households in some table cells (cross-classified by household size, income, presence of elderly, and receipt of public assistance) from which to draw participants. One of the reasons behind these discrepancies is that

Table of Contents

poor single adult households with children are underrepresented in the CPS (Ch. 3, p. 19).

3.2 Studies using sensitivity analyses

As part of the work of the Panel to Evaluate Microsimulation Models for Social Welfare Programs, Cohen et al. (1991) did an illustrative validation of TRIM2 for AFDC. Their goals included external validation and identifying which modules have large effects on model outputs. They used a sensitivity analysis for the latter, by using different versions of three modules. For two of these modules (modeling undercoverage for households with particular characteristics, and allocation of variables measured on an annual basis to months) they used one alternate version (two versions total), while for the third (static aging) they used three alternate versions.

Cohen et al. considered 16 different outputs for a program change that had actually occurred. They used a factorial experiment, analyzing the outputs with ANOVA. By taking the dependent variable to be the difference between the simulated output and the observed value, they were able simultaneously to do a sensitivity analysis and an external validation. They found that the method of aging contributed most to total variability, followed by adjusting the data for undercoverage. No one version of a module did clearly better than another in replicating observed totals, nor did any one combination of versions. They note that a more general model for validation would include using replicate datasets, with replication being a random effect and the model alternatives being fixed effects.

Atrostic and others at the Congressional Budget Office (CBO) considered sensitivity analyses in their development of a model of health services spending and payment (Atrostic 1994 and 1995; see Atrostic and Bilheimer, 1993, for a more detailed description of the model). The CBO model is based on the March 1993 CPS, combined with the 1987 National Medical Expenditure Survey (NMES). Since health care expenditures and health insurance premiums are not part of the CPS dataset, statistical matching is used to impute them from NMES. Match cells are created, and missing data for an individual is imputed based on the relevant match cell. In constructing the data file for the model, they considered multiple models of statistical matching to impute these variables.

Atrostic (1994) examined three different models for imputing expenditures. In two of these models, records within match cells are sorted by age, whereas in the third model they are also sorted by income within match cell. Since expenditures are highly skewed, outputs of interest include the mean, standard deviation, and percentiles of the expenditure distribution. Outputs of all three models replicated NMES distributions quite closely.

Atrostic also compared two models for jointly imputing expenditures and premiums. In one model, both were imputed based on the same set of variables. In the other, in addition to using the same set of variables, expenditures were imputed based on premiums and vice versa. The results of these two models were somewhat different. For example, the mean total health expenditure for people with expenditures differed by $200, and the mean premium differed by over $400.

In her 1995 paper, Atrostic describes strategies for choosing when to retain more than one version of a model specification in the development of the CBO model. Resource limitations prevented examining alternatives for every aspect of model specification in a classical experimental design. Instead, alternative model specifications were examined early on. In some cases the decision was made to use a single model specification for aspects that did not seem to contribute greatly to overall uncertainty, while retaining more than one alternative model specification for other aspects.

One set of model choices involved whether to impute premiums and expenditures jointly or separately. Their results suggested that they should be imputed jointly, but partly due to feasibility considerations, they decided to impute them separately. Another choice involved whether to match by family size or by size of health insurance unit, and it was decided to match using only health insurance unit size. For modeling what share of Federal taxes will be used to pay for health care, three different models were retained. Uncertainty about how state and local taxes are to be modeled led them to develop two alternatives. Atrostic also discusses several models of calibration, including applying ratio adjustments, or using a raking algorithm. In the CBO model they decided to apply only ratio adjustments for detailed spending categories, cross-classified by detailed financing categories.

By retaining several model versions written during the development of a microsimulation model, the CBO model will have a built-in mechanism for doing a sensitivity analysis for a number of different alternatives, for what we call model specification.

Table of Contents

3.3 Studies addressing stochastic simulation, calibration, or other sources of uncertainty

Kennickell and McManus (1994) use multiple imputation for item nonresponse in the data from the 1983 and 1989 panels of the Survey of Consumer Finances (SCF). The 1989 panel included panel cases based on the 1983 design as well as new cross-sectional cases (Heeringa et al., 1993, as cited in Kennickell and McManus, 1994). Missing data rates varied widely between variables, but for those households in the SCF in both 1983 and 1989, the percent with missing data in both years was fairly small. For missing data in the 1989 panel, software which imputes based on cross-sectional cases was modified to include conditioning on variables in the 1983 data.

Kennickell and McManus examined how much variability was added as a result of doing multiple imputation. They found that the coefficient of variation for the mean of a variable varied widely depending on what variable was being considered. In general, the coefficient of variation was largest for asset types held by fewer people, and smaller for aggregated measures such as total assets. For the imputation of 1989 total family income and components of financial assets changes, they examined how R^2 decreased as variables used in the imputation are dropped out. They compared this with and without using 1983 data. For some variables the inclusion of 1983 data gave a much larger R^2, and for most variables R^2 decreased substantially as the number of excluded variables increased.

Doyle and Farley (1994) examined the sensitivity of results of the AHSIM microsimulation model of health care reform to stochastic simulation. In the AHSIM model, health expenditures are imputed in a 2-part model (Doyle and Farley 1994, Farley and Doyle 1995). First, 10 probit models simulate the probability of having an expenditure for each of 10 types of services.

Table of Contents

mean expenditure. This example shows how calibration can reduce the overall uncertainty of model outputs. If more than one method of calibration is possible, the estimate of uncertainty does not include uncertainty about calibration method.

Farley and Doyle (1995) considered different methods for imputing expenditures and the effect of calibration on these methods. The standard method was described above (Doyle and Farley, 1994). Other methods included drawing the error terms for the log-linear model using two different non-parametric approaches, and another method imputed total expenditures directly, i.e. not categorized into type of expenditure. Since expenditures have a highly skewed distribution, the distribution of expenditures into increments and by percentiles was of interest, and several measures of comparing simulated results under different imputations with actual expenditures (from NMES) were used.

Farley and Doyle found that the distribution of simulated expenditures was similar to actual expenditures under each imputation. In the standard method, the distribution of expenditures was compressed and calibration to control totals did not fix this. One of their non-parametric methods, which preserves correlation among error terms for the same person among different services, performed the best overall. Although this method did not do well at replicating the distribution of expenditures at the top of the distribution, it most closely replicated the overall expenditure distribution. They also concluded that if total spending is of interest, the model which imputes total spending directly is preferable to the current method in which each expenditure is imputed separately.

Pudney and Sutherland (1994) incorporated three sources of uncertainty in the POLIMOD microsimulation model, a tax-benefit model for analyzing the distributional effects of policy changes. They incorporated uncertainty due to sampling, stochastic simulation, and parameter estimation. In addition they compared outputs under three specifications of the model for employment participation of women in the reform. They did not age their data, but pointed out that the aging process could contribute greatly to uncertainty.

Pudney and Sutherland consider a very radical policy change, in which all adults would be given an income, regardless of employment status, marital status, or income level. They expected that this type of reform would induce a large change in labor supply of women, and it is this aspect of behavior that they targeted for model parameter uncertainty. They allow for 3 possible states of employment: not working, working part time, and working full time. The predicted participation state for each woman in the base policy is generally simulated using a multinomial logit model.

Use of a resampling method to estimate sampling variability was not computationally feasible in their case. Instead they estimated sampling variability using asymptotic approximations.

Pudney and Sutherland compared outputs under three different models of employment participation for women in the reform. In the first two, participation in the base policy is simulated from the multinomial logit. In the first model, participation under the reform policy is simulated to be the same as under the base policy, whereas the second allows for second round effects. They compare this to a third model in which participation in the reform policy is the same as the observed state in the base policy.

For each of the three employment participation models, and for the base policy, the reform, and the difference between policies, Pudney and Sutherland calculated 11 model outputs each with a standard error. They assumed that the standard error for the difference between reforms was the same as the standard error for the reform policy totals. Consequently, they concluded for example that the coefficient of variation for total net payment in the reform is 10%, but is 160% for the change in net payments. Their results showed that sampling is the largest contributor to uncertainty for most outputs, with parameter uncertainty important for some outputs.

Pudney and Sutherland compared the relative importance of sampling, stochastic simulation and model parameter uncertainty, addressed in a way that we would consider each to be a random effect. However, without a Bayesian approach they cannot compare the relative importance of these uncertainties to what we would call the fixed effect of model specification, of employment participation for women in the reform. Instead they give point estimates and standard error estimates (based on the three random components) for the outputs separately for each employment participation model.

Table of Contents

4 Our treatment of uncertainties

Some types of uncertainties, such as sampling and stochastic simulation uncertainty, are due to a random effect and are most naturally summarized by a variance estimate. Other types of uncertainty, such as model specification and macro effects, can be called fixed effects. In this case particular levels are interpretable, and the interest may be in the output resulting from particular model specifications or particular realizations of a macro effect (Maxwell and Delaney, 1990, ch. 10). For example, one may be interested in the simulated model output if the unemployment rate was 2% less than the current rate.

Estimates of model parameters may be subject to sampling variability because the estimated model parameters are often based on supplementary studies based on samples. However, in some cases the analyst is interested in a particular value of the model parameter a priori. As an example, one may be interested in what the output would be if all eligible households participated in the Food Stamp Program (FSP).

Especially for fixed effects, the interest is often in model outcomes under particular model specifications, different macro assumptions, or different values of model parameters. An analysis of model sensitivity, or of how much the point estimate of the model output changes under alternative specifications, is called sensitivity analysis. In sensitivity analysis we are also interested in the variability of the estimated change in model outputs.

We use the term factor to denote any of the sources of uncertainty for which the programmer can vary the input levels or possible realizations. One way to estimate the contribution to variance of each error source is to run the model multiple times, varying the level of any factor, using an appropriate experimental design.

One simple and useful design is a full factorial design, in which the model is run under all possible combinations of the levels of the factors. For example, if the model uses factors A with n levels and B with m levels, there are nm total model runs (Cochran and Cox, 1950, ch. 5). In this design, A is said to be crossed with B. The effect of a particular factor can be estimated by averaging the output over the levels of all the other factors.

In a nested design, factor A is nested in factor B if each level of A is associated with only one level of B. For example, if a model uses m levels of factor A and n levels of factor B, then m/n distinct levels of A are run for each level of B, giving a total of m model runs (Snedecor and Cochran, 1980, ch. 16). In our work we use a combination of nested and crossed designs.

We use a Bayesian framework to interpret the importance of each error source. In a Bayesian framework, parameters are considered to be random. Before collecting data, we specify the distribution of each parameter by a prior probability distribution. The prior distribution represents the uncertainty about the parameter, and may reflect the range of values thought to be plausible by a single analyst, or it may reflect disagreements between different people. Ideally, the prior distribution should reflect the range of disagreement about the parameter value among people with expert opinion on the subject. Sometimes a flat prior is used, which gives equal prior probability to every possible value. A flat prior corresponds to a situation in which there is no prior information about the parameter.

After the prior distributions are specified, data is collected. In this case the data are the results of running the MATH-CPS model multiple times, under different levels of several factors. The prior distribution and the data are then combined using Bayes Theorem to give the posterior distribution for each parameter. Bayes Theorem says that the posterior distribution of the parameter, given the data, is proportional to the sampling distribution of the data times the prior distribution, $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$.

Using a Bayesian framework makes it possible to compare uncertainties about non-random elements such as the model specification, and random elements such as sampling, in a unified way. Without a Bayesian approach, point estimates of outputs can be compared under two different models or two different calibration methods (see for example Pudney and Sutherland 1994). However the degree to which model specification or calibration method contributes to the total uncertainty of the estimate of interest compared to random elements such as sampling, cannot be ascertained. Using a Bayesian approach makes it possible to consider both random and non-random elements as contributors to posterior variance and estimate the relative importance of each factor.

An estimate of the absolute size of each error source can be obtained from running the MATH-CPS model multiple times, outputting the point estimate of interest each time. The contributions of each factor to the variances of the model outputs can be summarized using analysis of variance, which can be easily applied in a full factorial experiment or a combined factorial and nested experiment.

For both fixed and random factors, the main effect of a factor can be estimated by the difference in the average point estimate under one level of the factor from the average point estimate, averaged over that factor. The sensitivity of the model to the factor of interest is estimated by the main

Table of Contents

effect. The main effect of model specification, for example, gives an estimate of the sensitivity of the point estimate to the model specified.

For fixed effects, the sum of the variances of the interaction of a particular fixed factor with each random factor (or the sum of variances of random factors nested within the fixed factor), gives an estimate of the variance associated with the main effect. For random factors, the variance of the main effect can be estimated from the variance component for the corresponding main effect.

We can also estimate the relative importance of each error source by looking at the contribution of each error source to posterior variance. The percentage of total posterior variance due to a particular factor gives an estimate of how important that factor is relative to other factors. The Bayesian framework makes it possible to compare the relative importance of both random and non-random factors.

Assuming linearity of the estimate with respect to the parameter, we can get a good estimate of the posterior parameter variance by using two values of the parameter based on the prior distribution: one at one standard deviation above the mean and the other at one standard deviation below the mean. Although it could be difficult to fully specify a prior distribution for a parameter, it might be possible to specify a prior mean and standard deviation, which is all that is needed to specify the two values.

There are two issues concerning calibration uncertainty: uncertainty about the totals or percentages to which the model is calibrated, and uncertainty about how to calibrate to these totals or percentages, i.e. uncertainty about calibration method. When a microsimulation model doesn't incorporation calibration, a particular total of interest will vary from run to run, due to sampling, stochastic and other sources of error.

When calibrating to a total of interest, the variability of this total due to sampling etc. is replaced with the prior uncertainty of the total. If the model is adjusted in such a way that the total for each run always exactly hits the same target, then effectively our prior uncertainty for the total is 0. If our prior uncertainty about the total is smaller than the sampling variability for the total, then we expect that the posterior variance of the total of interest would be reduced.

There could be many plausible methods to calibrate to a known total, and the posterior variance for the total should reflect this uncertainty. If calibration is done in one way only, then the posterior variance only includes variance within method of calibration, and the estimated uncertainty will be

Table of Contents

too small. By running a microsimulation model under at least two different methods of calibration, the posterior variance correctly includes variance between methods of calibration, as well as variance within calibration method.

We can use the information from the estimated variance components and their relative sizes as a guide to future simulations, future model changes, and future data collection. If the simulation variability is quite large relative to other components of posterior variance, this suggests doing more simulations with more random seeds in order to reduce the variability of model outputs. A large variance component for model specification uncertainty suggests more work should be done to specify the model as correctly as possible for use in future models. On the other hand, if sampling variability is large, perhaps an effort should be made to collect a larger sample for use in future models.

Table of Contents

5 The MATH-CPS model

The Micro Analysis of Transfer to Households (MATH-CPS) model is the most complex of the three food stamp microsimulation models written by Mathematica Policy Research (MPR). The MATH-CPS model is based on the Current Population Survey (CPS). In particular, the 1996 MATH-CPS model, with which we work, is based on the March 1993 CPS.

Like the MATH-SIPP model, the model of intermediate complexity written by MPR, the MATH-CPS model simulates eligibility and participation in the food stamp program. Unlike MATH-SIPP, the MATH-CPS model also simulates federal income and payroll taxes, which is not done in MATH-SIPP (MPR, 1994a).

The MATH-CPS data is organized and processed by household. As is done in the CPS, within each household, data is organized by family, and then by person within family. Each household and person has a weight associated with it, which is essentially the aged version of the March 1993 weights (MPR, 1994b).

5.1 Creation of the MATH-CPS database from the Current Population Survey

The CPS consists of a multistage stratified sample of non-institutionalized residents of the U.S. Each month approximately 71,000 households are selected to be interviewed of which 57,000 households containing approximately 112,000 people aged 15 or older are actually interviewed. (U.S. Department of Commerce, 1993a). Each household is interviewed monthly for four consecutive months, not interviewed for 8 months, then interviewed again for four consecutive months. During any given month, the CPS sample consists of 8 groups of households, each having entered the sample at a different time. These 8 groups are called rotation groups. Each rotation group is a probability sample of the population (U.S. Department of Commerce, 1978).

In the CPS, demographic data such as age, sex, and race are collected as are data on employment, such as occupation and hours worked. In the March CPS additional information is obtained, including work experience, employment history, and income, on an annual basis for the previous year.

For the March CPS, an additional 2500 Hispanic households are interviewed and supplementary data are collected for members of the Armed Forces who

Table of Contents

live on a military base or with their families in civilian housing units. The March CPS uses statistical matching and hot deck imputation to impute missing values. Statistical matching links each record which contains a missing value to a donor record without a missing value for the variable in question. Hot deck imputation imputes the missing value from the observed value in the donor record (U.S. Department of Commerce, 1993a).

The resulting data is representative of the United States as a whole, of each state, and of some other areas. The CPS is geared to producing national estimates, and for some purposes is inadequate to produce sufficiently precise estimates at the state level.

The CPS data is organized by household, then by family within household, and finally by person within family. Each type of unit has a weight associated with it. The calculation of weights for each person in the file is done in several stages. The basic weight is the inverse probability of selection. This weight is then adjusted for some special sampling situations and for non-interview of sampled households, within cluster by residence, race, and rotation group (U.S. Department of Commerce, 1978).

The next step in the calculation of weights uses two stages of ratio adjustments to adjust the sample to represent the known distribution of the population. The first-stage ratio adjustment is based on residence category, race, and region. The second-stage ratio adjustment is based on age-sex-race groups (U.S. Department of Commerce, 1978). Further adjustments to the weights are made to create the March CPS supplement weights, necessitated by the addition of households into the sample, and by the need for husband and wife to have the same weight.

Unlike the MATH-SIPP model, the MATH-CPS model ages the data, in this case from 1993 to 1996. The data is first aged demographically by readjusting person-level and household-level weights so that population counts by age (16 categories), sex, and race (4 categories) agree with projected counts. The projected counts are based on projections given in Table 2 in the Current Population Reports, P-25-1104 (U.S. Department of Commerce, 1993b). Then annual income variables are aged, by readjusting each type of wage according to the educational class of the individual. Finally, the annual poverty threshold is adjusted.

There are several problems in using the CPS data for food stamp microsimulation models. First, eligibility for the food stamp program and other public assistance programs depends on monthly rather than annual income, but the CPS only collects data on annual income. Secondly, some variables needed to determine eligibility for the food stamp program, such as financial and

Table of Contents

vehicular assets, and shelter, medical and deductible child care expenses, are not part of the CPS data (MPR, 1994a). Finally, AFDC, SSI and GA benefits, which affect eligibility for the food stamp program, are underreported in the CPS (US Dept of Commerce, 1992).

5.2 MATH-CPS model modules

The MATH-CPS model consists of a series of modules which may be run separately, or in conjunction with other modules. A very brief description of each MATH-CPS model module is given in Table 3 (MPR, 1994a). More details about each module may be found in Appendix 3.

The modules are ordered because some modules create variables which are then used in subsequent modules. For example, eligibility for the food stamp program is modeled in the FSTAMP module, and depends in part on whether or not the household was simulated to participate in AFDC, SSI, or GA. Participation in these three public assistance programs is simulated in the PBLAST module. Participation in public assistance programs depends on eligibility which is simulated in the PAPRAT module. The formation of public assistance filing units, needed by PAPRAT, is done in the UNIT7 module. These three modules in turn rely on earlier modules, such as ASSETS which imputes financial and vehicular assets, and ALLOY which allocates yearly earned income to months.

We are primarily interested in how a reform will affect changes in simulated participation and modeled benefits in the food stamp program (FSP). Food stamp eligibility and participation is simulated in the FSTAMP module, which can be run either under conditions for the baselaw or for the reform. This is the only MATH-CPS model which can be run under two different conditions. In a typical MATH-CPS model run, we run the FSTAMP module twice: first under the baselaw then under the reform.

Households which are eligible for the FSP under the baselaw, and which do not receive AFDC, are simulated to participate in the FSP under the baselaw with probabilities given in an array. These probabilities were derived in a series of steps. The initial probabilities are given in a 64-cell, 4-dimensional array, indexed by gross income relative to the poverty level (4 classes), unit size (4 classes), whether or not SSI or GA is received, and whether or not the household has elderly members. These initial probabilities are the number of participants (from the Food Stamp Integrated Quality Control Sample (IQCS) data) divided by the number of eligibles (from running the MATH-CPS model) (MPR, 1994a, ch 11). In some cells this initial probability is

Table of Contents

Table 3: MATH-CPS model modules

- MAKIO, MAKEHEAD, and MAKEMATH: converts CPS data into MATH-CPS database
- CODECPS: recodes CPS data to MATH-CPS variables
- DEFSTA: creates status variables
- DEMAGE: demographically ages the data to 1996
- CSWORK: tabulates unemployed people and calculates unemployment rate.
- ECONAGE: economically ages the data to 1996
- ALLOY: allocates yearly income to months
- ASSETS: imputes financial and vehicular assets
- CHILDEXP: imputes child care expenses
- MEDEXP: imputes medical expenses
- SHLTREXP: imputes shelter expenses
- UNIT7: creates filing units for public assistance (AFDC, SSI, GA)
- PBLAST: simulates means eligibility for public assistance
- PAPRAT: simulates participation in public assistance
- FSTAMP: simulates eligibility and participation in the FSP

Table of Contents

greater than one. The excess participants in these cells are allocated to other cells in order to keep key marginal distributions as close as possible to the estimated distributions. The key margins of interest include the distribution of earnings by unit size and by poverty level (separately), and the percentage of food stamp units with earnings, with elderly, and with children (aged 5-17).

This 64-cell array is then expanded into 2 additional dimensions: the ratio of the bonus value (simulated FSP benefit) to poverty line (with 6 classes), and whether or not the household reported receiving food stamps. In order to calibrate the results more closely to marginal totals, these participation rates are then multiplied by a number ranging from .5 to 2, depending on whether or not the unit has earners, and whether or not the unit receives SSI. (See Martini (1991) for more details about the MATH-CPS participation decision, and some suggestions for improvement.)

Simulated participation for an individual household under the reform depends on the household's simulated participation under the baselaw. Any household which participated in the FSP under the baselaw and is simulated to receive a larger benefit under the reform, is modeled to participate in the reform. Any household which was eligible for the FSP under the baselaw but didn't participate, and would receive a smaller benefit under the reform is modeled not to participate in the reform. There are only three household categories for which different participation decisions under the reform and the baselaw are possible:

1. Households which were not eligible for the FSP under the baselaw, but are eligible under the reform.
2. Households which were eligible for, but did not participate in the FSP under the baselaw, and are eligible for a larger benefit under the reform.
3. Household which were simulated to participate in the FSP under the baselaw, and would receive a smaller benefit under the reform.

In these three cases, the participation decision in the reform is stochastically simulated. Since participation in the reform is highly correlated with participation in the baselaw, the average difference between the two is kept small. (See Pudney and Sutherland (1994) for a different way of keeping this difference small.)

As mentioned earlier, the MATH-CPS database was created from running each module once in sequence, using one initial random seed (which we will

Table of Contents

refer to as the MPR seed). The random seed generates a vector of random numbers for each household. When we run a module using a changed set of conditions, such as using a different seed, the variables relevant to that module are in effect overwritten, whereas variables created under other modules remain unchanged. If an early module is not run in a particular simulation, subsequent modules use the values created from the initial MPR run of the module, which uses the MPR seed.

Table of Contents

6 Sources of uncertainty in the MATH-CPS model and our treatment of these uncertainties

6.1 Stochastic simulation

Imputation is used to create variables not available from the CPS data. This relies on use of random numbers. Random numbers are also used in generating individual error terms for regression equations, in simulating participation in public assistance and the FSP, and in calibration.

With the exception of the CODECPS module, only the last 9 MATH-CPS model modules (ALLOY through FSTAMP - see Table 3) make use of random numbers. We moved the part of CODECPS which uses random numbers to another module, and then only worked with the last 9 modules. A more detailed description of how random numbers are used in each module follows.

In ALLOY, random numbers are used to allocate the start date for employment income, for people who were employed only part of the year. Since the MATH-CPS model uses data from one month only, this would have an impact on whether or not an individual was simulated to be employed in the

Table of Contents

other earnings deductions are given) for the first four months that they are on AFDC and working.

In PAPRAT, participation in public assistance programs is simulated. Simulated participation probabilities are cell-based and depend on region, type of public assistance, and whether or not the household reported receiving welfare in the last calendar year. Any household which reported receiving welfare, and for which the target number of participants in the relevant cell is less than the number of eligible units for the public assistance program in question, is simulated to participate. Otherwise, the household is simulated to participate if a random number is less than the appropriate cell probability.

In the FSTAMP module, participation in the food stamp program under both the baselaw and the reform, for certain household categories, is stochastically simulated. For households which don't receive AFDC, the simulated participation decision was described in the section on MATH-CPS model modules. All households which are simulated to receive AFDC are simulated to receive food stamps.

A vector of random numbers is generated from a single random seed. The MATH-CPS model is set up so that for a given household and a given seed, the same random number will be used for a particular error term or simulated decision, regardless of the number of random numbers needed for that household prior to that point. We address stochastic simulation variability by running the model under several different seeds.

In the MATH-SIPP model, we used antithetic variables (Hammersley and Handscomb, 1964) to reduce stochastic variability. In that model, we found that although the antithetic strategy did reduce stochastic variability, sampling variability was very large relative to stochastic variability, so using antithetic variables did not give much reduction to total variance. Based on the MATH-SIPP work, and because of the additional programming effort required to implement the antithetic strategy, we decided not to use antithetic variables for our work with the MATH-CPS model. Use of antithetic variables might be worthwhile for future work with the MATH-CPS model however.

6.2 Sampling

Since the CPS is based on a sample of households, there is sampling uncertainty inherent in any point estimate produced from the MATH-CPS model.

Table of Contents

We used a grouped jackknife (Wolter, 1985) to estimate sampling variability. We estimated sampling variability under two different jackknife groupings: systematically leaving out every kth record with a different starting point in each run, or leaving out one rotation group in each run. If records are ordered randomly, the systematic jackknife leaves out random groups.

There are several advantages to using rotation group to define the jackknife replicate. Each rotation group is a probability sample of the population, and thus replicates the entire sampling scheme. Therefore the jackknife gives a proper estimate of variability taking into account the various kinds of stratification and clustering built into the CPS sampling scheme. When leaving out one rotation group, the remaining 7 rotation groups in the run are also a probability sample of the population.

By systematically dropping out every kth record, we are assured that if there is any stratification in the order of the records, we retain this stratification in the jackknife sample. One possible advantage to using the systematic jackknife is that we can obtain more degrees of freedom for estimates of sampling variability.

6.3 Model parameters

Model parameters are written into the code in many MATH-CPS modules. Demographic aging and economic aging of wages both use model parameters. Model parameters are used in the regression equations used to impute assets, child care expenses, medical expenses and shelter expenses. The modeled probabilities of participating in different types of public assistance programs (AFDC, GA, or SSI) can be thought of as model parameters. Finally, the array of probabilities of participating in the food stamp program, both under the base plan and under the reform (under certain conditions) are model parameters.

One way to address model parameter uncertainty is to pick two values of the parameter (or vector of parameters) from the prior distribution, ideally at $\pm$ one standard deviation from the mean. When we modeled parameter uncertainty in the MATH-SIPP model, we picked two arbitrary values for the parameter, for the purpose of illustrating this method. We made changes to the program in such a way that these values can easily be modified. We expected that someone with more subject knowledge about the parameters could easily modify these values accordingly.

We were interested in how sensitive the MATH-CPS model is to changing

Table of Contents

the parameters governing participation in the FSP under the baselaw. One disadvantage of the current set of parameters is that the MATH-CPS model does not simulate enough participants as compared to the target number of participants. We developed an alternate way to derive these parameters. We don't claim that our method is better, only that it is different, and that the difference in model outputs under these two methods would be of interest because the choice between the methods is arbitrary. We had planned to implement our changed set of parameters in the MATH-CPS model, but decided that other sources of variability were more important to address. Further description of this method can be found in Appendix A.5.

6.4 Model specification

Models are used, either implicitly or explicitly, in many different MATH-CPS modules. There may be a disagreement as to what model would be the most useful, and no model is likely to be entirely correct. A plausible alternative model might give a different outcome. Some examples of model specification in the MATH-CPS model follow.

Model specification uncertainty arises in the aging of wages, and in the implicit assumption that income from wages is distributed evenly across weeks worked, and that weeks worked are contiguous. Model specification also arises in the use of regression models to impute missing data, in particular data for assets, child care expenses, medical expenses, and shelter expenses. The assumption here is that a regression model is the appropriate model to specify the missing data, and that the correct variables are in the model.

The probability of participating in public assistance is modeled as a function of region, type of public assistance, and whether or not the household reported receiving welfare in the last calendar year. The probability of participating in the food stamp program is modeled as a function of several variables, as described in Section 5.2. In both cases, a particular parsimonious specification is assumed.

The assignment of weights to person and household records is based on implicit models, as described in Section 5.1. Likewise, the creation of food stamp units in a household relies on a model. If one person in a household receives food stamps, all other individuals in the household are modeled to receive food stamps as well, with the exception of individuals excluded by food stamp program rules (MPR, 1994a).

We now consider model specification uncertainty associated with the dis-

Table of Contents

tribution of income across months worked. The current MATH-CPS model allocates earned income evenly across all weeks worked. This may be realistic in many situations. However, for some individuals it is reasonable to think that income may have been earned at different rates during different weeks. Examples include people who work overtime on a seasonal basis, people who change jobs during the year, and school teachers who make extra money in the summer.

There are many ways in which income could be allocated unevenly across months worked. For example we could draw from a Dirichlet distribution to determine the percentage of earned income to allocate to each month worked. Alternatively we could draw from a Beta distribution, or from a Uniform[.25, .75] distribution, and allocate half the months to have earnings of the first random number times the mean, and the other half to have earnings of (2 - first random number) times the mean. A faster way of achieving a similar result is to work with the simulation month, and for the people affected, multiply their simulation month income by a Uniform[.5, 1.5] or Uniform[0, 2] random number.

An uneven allocation of earned income should have the effect of simulating some different people to be eligible for AFDC and for the food stamp program in any given month. If an individual's average earned income is just above the cutoff for eligibility for AFDC or the food stamp program, the individual would be income-ineligible for AFDC or the food stamp program under the current model. However the individual would have some chance of being income-eligible under the alternate model specification. As the modeled disparity between income received in different weeks increases, more individuals with average earned income just above the cutoff would have some chance to be income-eligible for AFDC or the food stamp program under the model of uneven income allocation. On the other hand, if an individual's average earned income is just below the cutoff and is eligible for AFDC or the FSP under the current model, there would be some chance that the individual would become income-ineligible under alternate model.

Our alternate method of income allocation only involved households below poverty level. For all such households, we multiplied the simulation month earned income by a Uniform[.5, 1.5] random number.

6.5 Calibration

Calibration is used in the MATH-CPS model in order to make many simulated totals come closer to projected totals (see Table 4 for some examples).

Table of Contents

Because a model must be specified in order to do this, calibration can be considered a type of model specification. Here we first describe the specifics of how and where calibration is used in the MATH-CPS model. Calibration of AFDC participants was thought to be particularly important, and we targeted this aspect of calibration for error analysis. After describing where calibration is used, we discuss an alternate calibration method for AFDC participants.

In addition to calibration of AFDC and food stamp participation, calibration is used in 4 modules: ASSETS, CHILDEXP, MEDEXP, and SHLTREXP. In each of these modules, a linear transformation is used to adjust the probability that a household will have the particular deduction (or asset), and to adjust the amount of the deduction (or asset).

In the ASSETS module, calibration is used for both financial and vehicular assets. Under the original, non-calibrated run of the MATH-CPS model, too few households were modeled to have financial assets, and the average financial asset was too large. The equations which simulate financial assets are now calibrated by randomly selecting 25% of households with income less than 130% of poverty and no assets to have asset holdings. Also the estimated amount of each household's assets is reduced by 20%. For vehicles, "negative adjustment factors" are applied to the equation that estimates number of vehicles, and the value of the first vehicle is reduced by 20% whereas the value of subsequent vehicles is reduced by 10% (MPR, 1994a).

In the CHILDEXP module, regression equations are used first to simulate whether or not a household has child care expenses, and then to model what the child care expense is, given that the household has such an expense. Uncalibrated simulations modeled too few households with child care expenses, and an average child care expense that was too large. To adjust, .5 is added to the simulated probability that a household has child care expenses, and the simulated child care expense is reduced by 5%.

Uncalibrated results for the MEDEXP modules showed that too many households had medical expense deductions, and that the average deduction was too large. The equation for imputing medical expenses is calibrated by multiplying the simulated expense by .50, and subtracting $50.

Similarly, uncalibrated results from the SHLTREXP module indicated that too many households had shelter expense deductions and that the average deduction was too large. The equations are calibrated by first multiplying shelter expenses by 1.2. Then $100 is subtracted from shelter expenses for rented households with income less than 50% of poverty, and $50 is subtracted from shelter expenses for rented households with income between 50

Table of Contents

Table 4: Some baselaw target quantities of interest in the MATH-CPS model, and the extent to which they are achieved.

Quantity	Target	MATH-CPS result	Fraction of target
Weighted number of units participating in FSP	10,864,000	9,204,000	.85
Weighted total benefits for FSP	2,029,742,000	1,679,354,000	.83
Weighted number of units participating in AFDC-Basic	4,630,000	4,268,000	.92
Weighted total benefits for AFDC-Basic	1,850,402,000	1,502,918,000	.81
Weighted number of units participating in AFDC-UP	307,000	231,000	.75
Total weighted benefits for AFDC-UP	180,236,000	104,172,000	.58
Percentage of FSP households with AFDC	40	37.0	.93
Percentage of FSP households with SSI	19	23.2	1.22
Percentage of FSP households with GA	8	6.0	.75
Percentage of FSP households with earners	20	23.1	1.16
Percentage of FSP households with elderly	15	21.7	1.45
Percentage of FSP households with children (5-17)	43	41.5	.97

Table of Contents

and 100% of poverty.

Calibration is also used to adjust the probabilities of participating in AFDC and in the food stamp program. We targeted participation in AFDC for development of an alternate calibration method. The MATH-CPS model simulates too few units to participate in the food stamp program and too few units to participate in AFDC (Table 4).

To come closer to the target number of food stamp participants, the MATH-CPS model takes some households which are eligible for the food stamp program but don't report receiving food stamps, and simulates them to participate. This is done by adjusting the simulated participation probabilities. For households which are eligible for food stamps and which don't receive AFDC, the simulated participation decision was described previously. The probabilities for participation were calibrated in the process of adjusting the participation probabilities for individual cells in the 64-cell matrix. This adjustment process was done to keep marginal totals as close as possible to projected totals. For households which are eligible for food stamps and which are simulated to receive AFDC, the participation probabilities are adjusted so that all such households are simulated to receive food stamps.

In order to calibrate AFDC participation to be closer to projected totals, the MATH-CPS model simulates participation for all families which are eligible for AFDC but don't report receiving them. Even though all eligible AFDC families are simulated to participate in AFDC, the simulated participation levels are substantially below the targeted participation levels.

The disadvantage to the calibration method used in the MATH-CPS model to increase participation in AFDC and the food stamp program is that the eligible non-reporting households, which are now modeled to participate, may be unlike the reporting households. An alternative calibration method is to readjust household weights, up-weighting people who report receiving food stamps, and/or up-weighting people who report receiving AFDC. We are interested in how much the estimated outcomes of interest are changed by an alternate, but also plausible, method.

We have addressed the calibration of AFDC totals by up-weighting AFDC-eligible families. As an alternative, we could have up-weighted only families which report receiving AFDC. This would have up-weighted fewer families by a larger amount. The justification for up-weighting these families could include the following:

1. There may have been a lower response rate in the CPS for AFDC recipients than for other households.

Table of Contents

2. There may have been a deficiency in the CPS sample which resulted in too few AFDC households in the sample.

3. Households may appear to be poorer when applying for AFDC than they appear to be in the CPS.

The MATH-CPS model processes two categories of AFDC separately: AFDC-Basic and AFDC-UP. Families which receive AFDC-Basic are primarily families with a single mother. AFDC-Basic families also include families consisting of a married couple in which one is disabled. AFDC-UP families are those containing two parents, but in which the principal earner is either unemployed or underemployed. The AFDC-Basic and AFDC-UP families are mutually exclusive.

For both AFDC categories, the MATH-CPS model simulates too few families and a total benefit which is too small, relative to the targets. The simulated benefits are farther from the target than the number of participating families, for both categories. (See Table 5, which differs slightly from Table 4 because they were produced using different seeds.) Our strategy to meet both the targeted number of families and the total benefits is to up-weight families which receive large benefits more than families which receive small benefits. We worked separately with AFDC-Basic and AFDC-UP families.

Our original method for re-weighting was to divide AFDC-Basic (and separately AFDC-UP) families into two groups: a low AFDC benefit group and a high AFDC benefit group. We sorted the relevant families by AFDC benefit, and for each possible cutoff between low and high benefit groups, we calculated the two weighting factors that would be needed in order to make both the total number of participants and the total benefits hit the targets. We calculated these factors by solving the following two equations for a and b:

$$\begin{aligned} aY_1 + bY_2 &= c_1(Y_1 + Y_2) \\ aB_1 + bB_2 &= c_1(B_1 + B_2) \end{aligned}$$

where

a = calibration weighting factor for the low benefit group

b = calibration weighting factor for the high benefit group
Y_1 = weighted number of households in the low benefit group
Y_2 = weighted number of households in the high benefit group
B_1 = weighted sum of AFDC benefits in the low benefit group
B_2 = weighted sum of AFDC benefits in the high benefit group
c_1 = the weighting factor needed to make the total number of AFDC participants match the target
c_2 = the weighting factor needed to make the total sum of AFDC benefits match the target

While any cutoff between low and high benefit group could be possible (assuming neither a nor b were negative or infinite), we were initially interested in not down-weighting either group, and in keeping both weighting factors fairly close to each other. When we found that grouping AFDC recipients into two classes was not particularly satisfactory, we extended this to three classes. We did this by picking a weighting factor for the low benefit group, then picking the cutoff between the medium and high benefit groups. We describe how we decided on the cutoffs in more detail in the Section 8.3. Zaslavsky (1988), discusses re-weighting to create new household records from Census data in order to calibrate to several targets, while minimizing the distance between individual weighting factors. His method takes into account more features of individual households and would have produced a less coarse calibration method than what was used here, but was too complicated to implement here.

6.6 Macro effects

We looked at uncertainty due to the macro effect of unemployment rate. Under the present MATH-CPS model, the simulated unemployment rate is based on the 1992 unemployment rate, which was 6.41% (using the MPR seed).

There are a number of ways to implement a changed unemployment rate. The simplest method, which is the least realistic, is to randomly "un-employ" say 2% of the individuals in the file. A second method which could be used either to decrease or increase the simulated unemployment rate, is to up-weight people who are employed during the simulation month and down-weight people who are unemployed in the simulation month (or vice versa if

Table of Contents

Table 5: Public Assistance Eligibility and Participation Summary in the MATH-CPS model, with families reporting AFDC listed separately from modeled participating families who didn't report receiving AFDC.

PUBLIC ASSISTANCE ELIGIBILITY AND PARTICIPATION SUMMARY

All states	Participants: Unweighted Units	Participants: Weighted Units	Participants: Weighted Benefits
Total AFDC-Basic (ADCTYPE 1-3,5-8)			
Reporters	1,733	2,926,019	$1,110,239,837
Non-reporters	787	1,318,124	$ 389,087,788
Total	2,520	4,244,143	$1,499,327,625
Target		4,630,000	1,850,402,000
fraction of target		0.92	0.81
AFDC-UP (ADCTYPE 4)			
Reporters	46	70,379	$ 35,520,374
Non-reporters	61	93,617	$ 37,660,307
Total	107	163,996	$ 73,180,682
Target		307,000	180,236,000
fraction of target		0.53	0.41

Table of Contents

we wanted to simulate a higher unemployment rate).

We implemented a third method, which could be used either to decrease or increase the unemployment rate. Under this method we find individuals who are unemployed for part of the year and employed for part of the year, but who are unemployed (or employed) during the simulation month. To decrease the unemployment rate we switch the simulation month for some individuals in the situation just described, so that the simulation month for these people are ones in which they were employed. The implicit model here is that changes in unemployment rate are mainly due to changes in the employment status of people who are occasionally employed.

When we switched the simulation month for all such people who were unemployed during the simulation month but employed during another month, using our seed file we got a 1.28% unemployment rate. We implemented a 4.27% unemployment rate by switching the simulation month for 40% of the people in the situation described.

We could have simulated a higher employment rate by switching the simulation month for people who were employed during the simulation month but unemployed during another month. When we switched the simulation month for all such people to a month in which they were unemployed, using our seed file got an unemployment rate of 15.76%. This represents the maximum unemployment rate that can be attained by this approach.

Since it is unlikely that anyone would want to simulate either an unemployment rate as low as 1.28% or as high as 15.76%, the number of people who are employed part of the year is not a limiting factor in using this methodology. If such an extreme unemployment rate was actually of interest, the assumption underlying this methodology – that a different unemployment rate was due to part-year employees – would not be a realistic assumption and a different method of simulating an extreme unemployment rate would be needed.

6.7 Food stamp unit size

The MATH-CPS model considers the food stamp unit to be the entire household, with the exception of people excluded based on food stamp regulations. In some cases the food stamp unit is actually a smaller subset of the household, and this error may have a large effect on the outcome of interest. Many alternate model specifications are possible. We have discussed an alternate model in which the person with the largest income in each household is taken

Table of Contents

out. Another model is to take out the oldest person in each household in one run of the MATH-CPS model, and to take out the youngest person in each household in another run. Although we discussed ways to address uncertainty in food stamp unit size, we did not implement this in this project.

Table of Contents

7 A method of quantifying uncertainties in the MATH-CPS model

7.1 Simulation experiment design

We have looked at a total of five sources of uncertainty in the MATH-CPS model:

1. Sampling, which we estimate by using a grouped jackknife, either leaving out random groups (systematic sampling) or leaving out one rotation group.

2. Stochastic, which we estimate by running the model under different random seeds.

3. Model uncertainty due to calibration of AFDC totals. We run the model under the current method of calibration as well as under a different method which up-weights households eligible for AFDC.

4. Model uncertainty due to method of allocating earned income to months worked. We run the model using the current method which allocates this income evenly, as well as under a method which allocates income unevenly.

5. The macro effect of unemployment. We run the model using the unemployment rate in the current model (6.4%), and under an unemployment rate of 4.3%.

Estimates of both the total posterior variance and the relative contribution of each source of error to posterior variance depend on:

1. What the outcome of interest is.

2. What reform is being considered.

3. Which MATH-CPS model modules are being run.

Some of the main outputs of interest are point estimates output by the MATH-CPS model in "Summary Comparisons of Impacts on Food Stamp Program" (see Table 6). The estimand of particular interest to users of the MATH-CPS model is the percent change in benefits under the reform. This is

Table of Contents

2. Write the expected mean squares in the simulation experiment ANOVA in terms of the variance components in the model.

3. Solve for the estimated variance components.

4. Calculate posterior variances for quantities of interest, using the estimated variance components.

In most of our experiments we used the design of seed nested in jackknife. We use a nested design rather than a design which would cross seed with jackknife for several reasons:

1. Based on earlier work with the MATH-SIPP model, we expected stochastic variability to be small relative to sampling variability.

2. We were not interested in the main effect of seed because seed has no meaning except in relation to a particular sample.

3. A nested design gives a more precise estimate of the variance component for seed because more seeds are used.

4. A nested design also gives a more precise estimate of overall means, since more seeds are used.

5. Under the assumption that stochastic simulation uncertainty is small, the estimate for sampling variability in a nested design should be sufficiently precise. (If this assumption is not met, a crossed design gives a better estimate of sampling variability.)

Our general notation for simulation results is y_{jscei} = an estimate of some quantity using

- jackknife replicate j ($j = 1, 2, \ldots J$)
- seed s ($s = 1, 2, \ldots S$)
- calibration level c ($c = -1, 1$)
- (un)employment level e ($e = -1, 1$)
- income allocation method i ($i = -1, 1$)

Table of Contents

We may use a subset of these subscripts, omitting those which are not varied in a particular experiment.

The percent change in benefits can be written as $\frac{\sum_h w_h y_{hjscei} - \sum_h w_h y'_{hjscei}}{\sum_h w_h y_{hjscei}}$, where h indexes households, y_{hjscei} = the benefit simulated to be received by household h under the base plan, and y'_{hjscei} = the benefit simulated to be received by household h under the reform plan. Here we assume weights are fixed, although they may vary in different runs which would necessitate additional subscripts.

Our output from a jackknife replicate or rotation group can be written as $y_{jscei} = f(c'\bar{y}) = \frac{\bar{y}_1}{\bar{y}_2} = \frac{\sum_{h \in j} w_h y_{hjscei} - \sum_{h \in j} w_h y'_{hjscei}}{\sum_{h \in j} w_h y_{hjscei}}$.

7.2 A model for the influence of observations

We develop a model for the mean $\bar{y}$, which applies directly to linear functions of $\bar{\mathbf{y}}$, $\mathbf{c}'\bar{\mathbf{y}}$. Through the Taylor linearization, the model applies to a nonlinear function of $\bar{\mathbf{y}}$, $f(\bar{\mathbf{y}})$. The Taylor expansion tells us that the jackknife variance estimator which gives an approximately unbiased estimate of the variance of a mean, also gives an approximately unbiased estimate for a function of a mean, such as the percent change in benefits which is a ratio of means.

In general,

$$f(\bar{\mathbf{y}}) \approx f(\mathbf{y}_0) + \left(\frac{\partial f}{\partial \mathbf{y}}\right)'\bigg|_{\mathbf{y}_0} (\bar{\mathbf{y}} - \mathbf{y}_0) = c_0 + \mathbf{c}'\bar{\mathbf{y}}$$

In the case of a ratio, $y_{jscei} = f(\mathbf{c}'\bar{\mathbf{y}})$, $\bar{y}_1 = \sum_h w_h y_{hjscei} - \sum_h w_h y'_{hjscei}$ and $\bar{y}_2 = \sum_h w_h y_{hjscei}$. If we expand this using the Taylor expansion we have

$y_{jscei} = \frac{\bar{y}_1}{\bar{y}_2} = f(\bar{y}_1, \bar{y}_2) \approx c_0 + c_1\bar{y}_1 + c_2\bar{y}_2$.

For the experiments in which we model all five sources of uncertainty (which we will refer to as the 5-factor case), we use the following decomposition for y_{jscei}:

Table of Contents

$$
\begin{aligned}
y_{jscei} = \tfrac{J}{n(J-1)} \textstyle\sum_{h\in j} \quad & (\beta + \beta_{H(h)} + \beta_{JHS(jhs)} \\
& +c(\beta_C + \beta_{HC(hc)} + \beta_{JHSC(jhsc)}) \\
& +e(\beta_E + \beta_{HE(he)} + \beta_{JHSE(jhse)}) \\
& +i(\beta_I + \beta_{HI(hi)} + \beta_{JHSI(jhsi)}) \\
& +ce(\beta_{CE} + \beta_{HCE(hce)} + \beta_{JHSCE(jhsce)}) \\
& +ci(\beta_{CI} + \beta_{HCI(hci)} + \beta_{JHSCI(jhsci)}) \\
& +ei(\beta_{EI} + \beta_{HEI(hei)} + \beta_{JHSEI(jhsei)}) \\
& +cei(\beta_{CEI} + \beta_{HCEI(hcei)} + \beta_{JHSCEI(jhscei)}))
\end{aligned}
$$

n is the number of records in the dataset (n=58971), and J is the number of jackknife replicates. Recall that c indexes calibration, e indexes unemployment rate and i indexes method of income allocation, where c, e, i are 1 to indicate the new model, and -1 to indicate the standard model.

The notation is simpler for the experiments for which we considered uncertainty due to sampling, stochastic simulation, and one fixed factor such as calibration method (which we will refer to as the 3-factor case). In this case,

$$y_{jsc} = \tfrac{J}{n(J-1)} \textstyle\sum_{h\in j}(\beta + \beta_{H(h)} + \beta_{JHS(jhs)} + c(\beta_C + \beta_{HC(hc)} + \beta_{JHSC(jhsc)})).$$

For experiments involving just sampling and stochastic uncertainty (the 2-factor case),

$$y_{js} = \tfrac{J}{n(J-1)} \textstyle\sum_{h\in j}(\beta + \beta_{H(h)} + \beta_{JHS(jhs)})$$

The grand mean β, the calibration effect β_C, the unemployment effect β_E, the income effect β_I, and the 2-way and 3-way interactions of calibration, unemployment and income are fixed effects of interest. All other effects are modeled to have mean zero since they are defined as deviations from the main effect across the entire population of households (for household, i.e. sampling effects) or the entire distribution of random numbers (for seed, i.e. stochastic simulation effects).

We define variances for each of the random effects, as shown below:

Table of Contents

$\beta_{H(h)}$	σ_h^2	main household effect
$\beta_{JHS(jhs)}$	σ_s^2	seed nested in jackknife
$\beta_{HC(hc)}$	σ_{hc}^2	calibration $\times$ household
$\beta_{JHSC(jhsc)}$	σ_{sc}^2	calibration $\times$ seed
$\beta_{HE(he)}$	σ_{he}^2	unemployment $\times$ household
$\beta_{JHSE(jhse)}$	σ_{se}^2	unemployment $\times$ seed
$\beta_{HI(hi)}$	σ_{hi}^2	income allocation $\times$ household
$\beta_{JHSI(jhsi)}$	σ_{si}^2	income allocation $\times$ seed
$\beta_{HCE(hce)}$	σ_{hce}^2	calibration $\times$ unemployment $\times$ household
$\beta_{JHSCE(jhsce)}$	σ_{sce}^2	calibration $\times$ unemployment $\times$ seed
$\beta_{HCI(hci)}$	σ_{hci}^2	calibration $\times$ income allocation $\times$ household
$\beta_{JHSCI(jhsci)}$	σ_{sci}^2	calibration $\times$ income allocation $\times$ seed
$\beta_{HEI(hei)}$	σ_{hei}^2	unemployment $\times$ income allocation $\times$ household
$\beta_{JHSEI(jhsei)}$	σ_{sei}^2	unemployment $\times$ income allocation $\times$ seed
$\beta_{HCEI(hcei)}$	σ_{hcei}^2	calibration $\times$ unemployment $\times$ income allocation $\times$ household
$\beta_{JHSCEI(jhscei)}$	σ_{scei}^2	calibration $\times$ unemployment $\times$ income allocation $\times$ seed

σ_h^2 could be interpreted as a between-household variance component, incorporating any design effects that are captured by the jackknife replication scheme. σ_s^2 could be interpreted as the mean stochastic variability for household. Each of the variance components for the interactions can be interpreted as the variance for the contrasts associated with each fixed main effect and interaction.

For convenience of notation in calculating expected mean squares for the simulation experiment outputs, we define "variances" for the fixed effects (σ_c^2 for the calibration effect, $\sigma_e^2, \sigma_i^2, \sigma_{ce}^2, \sigma_{ci}^2, \sigma_{ei}^2$ and σ_{cei}^2 for the other fixed effects), but since these are fixed effects, these latter "variances" don't have any meaning in our model for the outcomes of a simulation experiment.

There are at least three questions we might wish to answer from the MATH-CPS experiment outputs:

1. What is the estimate of and posterior variance of the grand mean, $V(\beta \mid \hat{\beta})$?

2. What is the estimate of and posterior variance of each of the fixed effects, $V(\beta_C \mid \hat{\beta}_C)$, $V(\beta_E \mid \hat{\beta}_E)$, $V(\beta_I \mid \hat{\beta}_I)$, $V(\beta_{CE} \mid \hat{\beta}_{CE})$, $V(\beta_{CI} \mid \hat{\beta}_{CI})$, $V(\beta_{EI} \mid \hat{\beta}_{EI})$, and $V(\beta_{CEI} \mid \hat{\beta}_{CEI})$?

3. What is the estimate of and posterior variance of the true overall effect? For the 2-factor case, the posterior variance of the true overall effect is $V(\beta \mid \hat{\beta})$, which is just the posterior variance of the grand mean.

Table of Contents

For the 3-factor case, the posterior variance of the true overall effect is $V(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$, and for the 5-factor case, this quantity is

$V(\beta + c\beta_C + e\beta_E + i\beta_I + ce\beta_{CE} + ci\beta_{CI} + ei\beta_{CI} + cei\beta_{CEI} \mid$
$\hat{\beta}, \hat{\beta}_C, \hat{\beta}_E, \hat{\beta}_I, \hat{\beta}_{CE}, \hat{\beta}_{CI}, \hat{\beta}_{EI}, \hat{\beta}_{CEI})$

Recall that c, e, and i indicate indicates which values of β_C, β_E, and β_I are correct respectively.

We are interested in each of these quantities under what we call a "full run" of the MATH-CPS model, as well as under a "single run" of the model. A full run corresponds to running the model many times as we do here, using a full factorial $\times$ nested design. We would also like to know how the total posterior variance, and the relative contributions of each source of uncertainty to posterior variance, would change if the sampling and stochastic simulation parts of the model were to be run only once, but the model would be run under each level of the fixed effects . In this case, we would have to use estimates of the variance components from a full run. A single run would give a larger posterior variance than a full run, since it would be impossible to average estimates of the sampling and stochastic elements over multiple runs. By running the model more times, the variance components for stochastic simulation error and stochastic error crossed with fixed effects are reduced, while the variance components for sampling, fixed effects, and sampling crossed with fixed effects are not. Running the model multiple times decreases the total posterior error, and decreases the relative importance of stochastic error.

7.3 Estimation of quantities of interest

Having written a variance components model for the population and defined quantities of interest, we now discuss how to estimate these quantities from a full run. We can calculate the expected mean square for each factor in the model. To do so it is helpful to first calculate each applicable mean of y, which we show below for the 3-factor case. We use a dot subscript to indicate an average over the corresponding factor.

$$y_{jsc} = \beta + c\beta_C + \frac{J}{n(J-1)} \sum_{h \in j} (\beta_{H(h)} + \beta_{JHS(jhs)} + c(\beta_{HC(hc)} + \beta_{JHSC(jhsc)}))$$

$$y_{j.c} = \beta + c\beta_C + \frac{J}{n(J-1)} \sum_{h \in j} (\beta_{H(h)} + c\beta_{HC(hc)} + \frac{1}{S} \sum_{s \in j} (\beta_{JHS(jhs)} + c\beta_{JHSC(jhsc)}))$$

Table of Contents

$$
\begin{aligned}
y_{..c} &= \beta + c\beta_C + \frac{1}{n}\sum_h(\beta_{H(h)} + c\beta_{HC(hc)}) \\
&\quad + \frac{1}{nS(J-1)}\sum_j\sum_{h\in j}\sum_{s\in j}(\beta_{JHS(jhs)} + c\beta_{JHSC(jhsc)}) \\
y_{js.} &= \beta + \frac{J}{n(J-1)}\sum_{h\in j}(\beta_{H(h)} + \beta_{JHS(jhs)}) \\
y_{j..} &= \beta + \frac{J}{nS(J-1)}\sum_{h\in j}\sum_{s\in j}(\beta_{H(h)} + \beta_{JHS(jhs)}) \\
y_{...} &= \beta + \frac{1}{n}\sum_h \beta_{H(h)} + \frac{1}{nS(J-1)}\sum_j\sum_{h\in j}\sum_{s\in j}\beta_{JHS(jhs)}
\end{aligned}
$$

Calculation of expected mean squares for the ANOVA are not presented here, but we present the results for the 2-factor, 3-factor, and 5-factor case in Appendix A.2. Once we know the expected mean squares, we can get estimates for each variance component of interest using the methods of moments (MOM) estimators. The MOM estimator equates expected and observed values, allowing us to solve algebraically for each variance component. Each variance component estimate is a linear function of the ANOVA mean squares and constants.

For the 2-factor case, Tables 28, 29, and 30 give the ANOVA sums of squares, expected mean squares, and estimators of variance components respectively for the nested design. We also give the same three results for a crossed design, in Tables 31, 32, and 33.

For the 3-factor case, Tables 34, 35 and 36 give the sums of squares, expected means squares, and variance estimators respectively. For the 5-factor case we show the sums of squares in Table 37, the expected mean squares in Table 38, and the estimators of variance components in Table 39.

The estimator for the grand mean $\hat{\beta}$ is the average over all conditions: $y_{..}$ for the 2-factor case, $y_{...}$ for the 3-factor case, and $y_{.....}$ for the 5-factor case. In each case,

$$
\begin{aligned}
\hat{\beta} &= \beta + \frac{1}{n}\sum_h \beta_{H(h)} + \frac{1}{n(J-1)S}\sum_j\sum_{s\in j}\sum_{h\in j}\beta_{JHS(jhs)} \\
&= \beta + \beta_{H(.)} + \beta_{JHS(...)}
\end{aligned}
$$

Table of Contents

with sampling variance

$$V(\hat{\beta} \mid \beta) = \frac{\sigma_h^2}{n} + \frac{\sigma_s^2}{nS(J-1)}.$$

The estimator of the calibration effect, $\hat{\beta}_C$, is half the difference of the average effect when one calibration method is used from the average effect when the other calibration method is used, or the difference of the average under one level of calibration from the grand mean. For the 3-factor case, $\hat{\beta}_C = y_{..1} - y_{...}$. For the 5-factor case, $\hat{\beta}_C = y_{..1..} - y_{.....}$. In both cases,

$$\begin{aligned} \hat{\beta}_C &= \beta_C + \frac{1}{n}\sum_h \beta_{HC(hc)} + \frac{1}{nS(J-1)}\sum_j \sum_{s\in j} \sum_{h\in j} \beta_{JHSC(jhsc)} \\ &= \beta_C + \beta_{HC(.c)} + \beta_{JHSC(...c)} \end{aligned}$$

with variance

$$V(\hat{\beta}_C \mid \beta_C) = \frac{\sigma_{hc}^2}{n} + \frac{\sigma_{sc}^2}{nS(J-1)}.$$

For the 5-factor case, the estimators and variances for each of the main, 2-way interactions and 3-way interactions are given in Table 7.

We put flat and independent priors on β, β_C, β_E, β_I, β_{CE}, β_{CI}, β_{EI}, and β_{CEI}. With a flat prior and under the assumption that the estimators are unbiased, the posterior means of each of these quantities equal their sample means:

$$\begin{aligned} E(\beta \mid \hat{\beta}) &= E(\hat{\beta} \mid \beta) \\ E(\beta_C \mid \hat{\beta}_C) &= E(\hat{\beta}_C \mid \beta_C) \\ E(\beta_E \mid \hat{\beta}_E) &= E(\hat{\beta}_E \mid \beta_E) \\ E(\beta_I \mid \hat{\beta}_I) &= E(\hat{\beta}_I \mid \beta_I) \\ E(\beta_{CE} \mid \hat{\beta}_{CE}) &= E(\hat{\beta}_{CE} \mid \beta_{CE}) \\ E(\beta_{CI} \mid \hat{\beta}_{CI}) &= E(\hat{\beta}_{CI} \mid \beta_{CI}) \\ E(\beta_{EI} \mid \hat{\beta}_{EI}) &= E(\hat{\beta}_{EI} \mid \beta_{EI}) \\ E(\beta_{CEI} \mid \hat{\beta}_{CEI}) &= E(\hat{\beta}_{CEI} \mid \beta_{CEI}) \end{aligned}$$

Also with a flat prior, the posterior variances of β and of each fixed effect equal their sampling variances:

Table of Contents

Table 7: Estimators and variances for calibration, unemployment, and income allocation effects and their 2-way and 3-way interactions

Effect	Estimator	Value	Posterior variance
β_C	$\hat{\beta}_C = y_{..c..} - y_{.....}$	$\beta_C + \beta_{HC(.c)}$ $+\beta_{JHSC(...c)}$	$V(\hat{\beta}_C \mid \beta_C)$ $= \frac{\sigma^2_{hc}}{n} + \frac{\sigma^2_{sc}}{nS(J-1)}$
β_E	$\hat{\beta}_E = y_{...e.} - y_{.....}$	$\beta_E + \beta_{HE(.e)}$ $\beta_{JHSE(...e)}$	$V(\hat{\beta}_E \mid \beta_E)$ $= \frac{\sigma^2_{he}}{n} + \frac{\sigma^2_{se}}{nS(J-1)}$
β_I	$\hat{\beta}_I = y_{....i} - y_{.....}$	$\beta_I + \beta_{HI(.i)}$ $+\beta_{JHSI(...i)}$	$V(\hat{\beta}_I \mid \beta_I)$ $= \frac{\sigma^2_{hi}}{n} + \frac{\sigma^2_{si}}{nS(J-1)}$
β_{CE}	$\hat{\beta}_{CE} = y_{..ce.} - \hat{y}_{..ce.}$ $= y_{..ce.} - y_{..c..} - y_{...e.} + y_{.....}$	$= \beta_{CE} + \beta_{HCE(.ce)}$ $+\beta_{JHSCE(...ce)}$	$V(\hat{\beta}_{CE} \mid \beta_{CE})$ $= \frac{\sigma^2_{hce}}{n} + \frac{\sigma^2_{sce}}{nS(J-1)}$
β_{CI}	$\hat{\beta}_{CI} = y_{..c.i} - \hat{y}_{..c.i}$ $= y_{..c.i} - y_{..c..} - y_{....i} + y_{.....}$	$\beta_{CI} + \beta_{HCI(.ci)}$ $+\beta_{JHSCI(...ci)}$	$V(\hat{\beta}_{CI} \mid \beta_{CI})$ $= \frac{\sigma^2_{hci}}{n} + \frac{\sigma^2_{sci}}{nS(J-1)}$
β_{EI}	$\hat{\beta}_{EI} = y_{...ei} - \hat{y}_{...ei}$ $= y_{...ei} - y_{...e.} - y_{....i} + y_{.....}$	$\beta_{EI} + \beta_{HEI(.ei)}$ $+\beta_{JHSEI(...ei)}$	$V(\hat{\beta}_{EI} \mid \beta_{EI})$ $= \frac{\sigma^2_{hei}}{n} + \frac{\sigma^2_{sei}}{nS(J-1)}$
β_{CEI}	$\hat{\beta}_{CEI} = y_{..cei} - \hat{y}_{..cei} =$ $y_{..cei} - y_{..ce.} - y_{..c.i} - y_{...ei}$ $+y_{..c..} + y_{...e.} + y_{....i} - y_{.....}$	$\beta_{CEI} + \beta_{HCEI(.cei)}$ $+\beta_{JHSCEI(...cei)}$	$V(\hat{\beta}_{CEI} \mid \beta_{CEI})$ $= \frac{\sigma^2_{hcei}}{n} + \frac{\sigma^2_{scei}}{nS(J-1)}$

Table of Contents

$$
\begin{aligned}
V(\beta \mid \hat{\beta}) &= V(\hat{\beta} \mid \beta) \\
V(\beta_C \mid \hat{\beta}_C) &= V(\hat{\beta}_C \mid \beta_C) \\
V(\beta_E \mid \hat{\beta}_E) &= V(\hat{\beta}_E \mid \beta_E) \\
V(\beta_I \mid \hat{\beta}_I) &= V(\hat{\beta}_I \mid \beta_I) \\
V(\beta_{CE} \mid \hat{\beta}_{CE}) &= V(\hat{\beta}_{CE} \mid \beta_{CE}) \\
V(\beta_{CI} \mid \hat{\beta}_{CI}) &= V(\hat{\beta}_{CI} \mid \beta_{CI}) \\
V(\beta_{EI} \mid \hat{\beta}_{EI}) &= V(\hat{\beta}_{EI} \mid \beta_{EI}) \\
V(\beta_{CEI} \mid \hat{\beta}_{CEI}) &= V(\hat{\beta}_{CEI} \mid \beta_{CEI})
\end{aligned}
$$

We put flat priors on c, e, and i which represents our uncertainty about which method of calibration is correct, which unemployment rate is correct, and which method of income allocation is correct:

$$
c = \begin{cases} 1 & \text{with probability .5} \\ -1 & \text{with probability .5} \end{cases}
$$

$$
e = \begin{cases} 1 & \text{with probability .5} \\ -1 & \text{with probability .5} \end{cases}
$$

$$
i = \begin{cases} 1 & \text{with probability .5} \\ -1 & \text{with probability .5} \end{cases}
$$

Note that the prior means of c, e, and i are all 0, while the prior variances are all 1. We want to calculate the posterior variance of the true overall effect, using the correct calibration method, the correct unemployment rate, and the correct income allocation method, conditional on our estimates. For the 2-factor case, this quantity is just $V(\beta \mid \hat{\beta})$. For the 3-factor case, the posterior variance of the true overall effect given our estimates is $V(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$, and for the 5-factor case, this quantity is $V(\beta + c\beta_C + e\beta_E + i\beta_I + ce\beta_{CE} + ci\beta_{CI} + ei\beta_{EI} + cei\beta_{CEI} \mid \hat{\beta}, \hat{\beta}_C, \hat{\beta}_E, \hat{\beta}_I, \hat{\beta}_{CE}, \hat{\beta}_{CI}, \hat{\beta}_{EI}, \hat{\beta}_{CEI})$

Below we show this calculations for the 3-factor case. The calculation for the 5-factor cases is similar. Note that for each of the fixed factors (calibration, unemployment rate, and income allocation method) the posterior variance incorporates the variability of the point estimates under the different levels of each fixed effect (i.e. the square of the difference in point estimates under one level from the grand mean), as well as uncertainty about which levels are correct.

Table of Contents

$$
\begin{aligned}
&V(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C) \\
&= V\left\{E(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C, c) \mid \hat{\beta}, \hat{\beta}_C\right\} + E\left\{V(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C, c) \mid \hat{\beta}, \hat{\beta}_C\right\} \\
&= V(\hat{\beta} + c\hat{\beta}_C \mid \hat{\beta}, \hat{\beta}_C) + E\left\{V(\hat{\beta} \mid \beta) + c^2 V(\hat{\beta}_C \mid \beta_C))\right\} \\
&= \hat{\beta}_C{}^2 V(c) + V(\hat{\beta} \mid \beta) + V(\hat{\beta}_C \mid \beta_C) E(c^2) \\
&= \hat{\beta}_C{}^2 + V(\hat{\beta} \mid \beta) + V(\hat{\beta}_C \mid \beta_C) \\
&= \hat{\beta}_C{}^2 + \left(\frac{\sigma_h^2}{n} + \frac{\sigma_s^2}{nS(J-1)}\right) + \left(\frac{\sigma_{hc}^2}{n} + \frac{\sigma_{sc}^2}{nS(J-1)}\right)
\end{aligned}
$$

We can see that the posterior variance of the true overall effect has five parts: uncertainty due to calibration, sampling, stochastic simulation, calibration $\times$ sampling, and calibration $\times$ stochastic, respectively. For a single run of the MATH-CPS model, the posterior variance is:

$$
V(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C) = \hat{\beta}_C{}^2 + \left(\frac{\sigma_h^2}{n} + \frac{\sigma_s^2}{n}\right) + \left(\frac{\sigma_{hc}^2}{n} + \frac{\sigma_{sc}^2}{n}\right)
$$

Table of Contents

8 Results from the MATH-CPS model

We present results first for experiments using sampling and stochastic simulation uncertainty, using a varying number of MATH-CPS modules. Next we discuss experiments also involving just sampling and stochastic simulation uncertainty, but using all 10 MATH-CPS modules. In these experiments we try to pinpoint the absolute and relative size of these two sources of uncertainty for the main reform we considered. Next we present experiments which in addition involve uncertainty due to calibration. Finally we give results for experiments involving all five sources of uncertainty, in which we compare different reform plans.

For the first three categories of experiments (i.e. all except those involving all five sources of uncertainty), we use the same reform plan. The reform plan we used differs from the baselaw in three ways:

1. The asset limit is increased to $5000 (versus $3000 for households with elderly, and $2000 for households without elderly in the baselaw).
2. The dependent care cap for the continental US is removed.
3. The shelter cap is reduced by 50%.

Increasing the asset limit should result in more participants. Removing the dependent care cap should give some gainers (people who are better off under the reform), and reducing the shelter cap should give some losers (people who are worse off under the reform).

8.1 Effect of number of modules on stochastic variability

We did five experiments which incorporated stochastic and sampling variability. In all five experiments we used a jackknife with 20 jackknife replicates, and 4 seeds nested in each jackknife (Table 8).

In our first experiment we just worked with the FSTAMP module under the reform. In this case, the percent change in benefits has uncertainty only due to the total benefits under the reform, since the total benefits in the baselaw remains constant in each run, in this case. We found (Tables 40 and 8) that sampling variability was extremely important compared to stochastic simulation uncertainty, with sampling accounting for 99.889% of the total posterior

Table of Contents

Table 8: Summary results for experiments varying the number of MATH-CPS model modules run. Experiments incorporate sampling and stochastic simulation uncertainty.

Modules	FSTAMP baselaw	FSTAMP baselaw and reform	8 modules	9 modules	10 modules
$\hat{\beta}$	9.827	9.738	9.932	9.234	9.215
$\sqrt{\sigma_h^2/n}$	0.805	.699	.479	0	0
$\sqrt{\sigma_s^2/n}$	0.233	.218	.233	.714	.709
SE(β | $\hat{\beta}$)					
single run:	.838	.732	.528	.714	.709
full run:	.805	.700	.479	.082	.081
% var due to seed					
single run:	7.763	8.894	17.794	100.0	100.0
full run:	0.111	0.128	0.284	100.0	100.0

variance of the mean. This is not too surprising, because the simulated participation decisions for the reform for several types of households do not rely on random numbers.

In the next experiment, we ran the FSTAMP module twice, first under the baselaw and then under the reform. We expected stochastic simulation to be a more important contributor to posterior variance than it had been in the previous experiment because the simulated participation decision under the baselaw for many households relies on random numbers. It turned out that the estimate of the variance component for seed was similar, and sampling variability accounted for 99.872% of the total variability (Tables 41 and 8). Although the baselaw participation rate is more dependent on random seeds, the difference in reform versus baselaw participation rate appears not to depend much on the simulated baselaw values.

We then ran the MATH-CPS model using all modules which contain stochastic variability, from ALLOY through the 2 FSTAMP modules (10 modules in all, since FSTAMP was run twice). Not too surprisingly, the estimate for the variance component for stochastic simulation was much larger (Tables 44 and 8).

The method of moments (MOM) estimators for the variance components were described in Section 7.3. The MOM estimator for σ_h^2 is proportional to MS(jackknife) minus MS(seed). Since each mean square is estimated with uncertainty, it is possible to get a negative MOM estimate for σ_h^2, and this

Table of Contents

occurred in the experiment just described. When a variance estimate is negative, the estimate is typically truncated to zero since variances must be non-negative. Because we continued to work with the same jackknife replicates and seeds throughout many of the remaining experiments, we continued to get negative estimates for σ_h^2. In the next section we discuss a better estimator of σ_h^2.

In order to pinpoint the module which most of the stochastic variability was coming from, we did 2 additional experiments. In the fourth experiment, we ran 8 modules, omitting only ASSETS and ALLOY. The result was that sampling variability was very large relative to stochastic simulation uncertainty (Tables 42 and 8). In the fifth experiment we ran 9 modules, omitting only ALLOY. The result of this experiment was that sampling variability was tiny compared to stochastic uncertainty (Tables 43 and 8). Thus it appears that the ASSETS module is the primary source of stochastic simulation uncertainty which is reasonable since the reform we considered involved a change in asset limits. For a reform which doesn't involve assets, the ASSETS module may not be such an important source of stochastic simulation uncertainty.

The ASSETS module incorporates stochastic simulation uncertainty for 2 types of operations: adding an error term to each of 6 regression equations, and calibration of financial assets, by simulating some households to have asset holdings. It would be relatively easy to narrow down the precise part of the ASSETS module which contributes the most to stochastic variability, by using different random seeds for parts of the ASSETS module only.

Table 8 shows an unexpected result which needs some explanation. As more modules are run, $\sqrt{\hat{\sigma}_h^2/n}$ decreases. One possible explanation for this follows.

Recall that if a module is not run in an experiment, this is equivalent to running that module under one seed, the MPR seed. Suppose a module contributes a lot to stochastic variability, as is the case for the ASSETS module. If this module is run under only one seed, then the variability we see resulting from that module is incorporated into the estimate of σ_h^2. As an extreme case, suppose the result of that module would be the same for each household if there was no stochastic element, so there is no sampling variability at all. Now suppose a stochastic element is added to each household, such as imputing a random subset of households to receive \$10 in asset income or adding an error term to a regression equation, but that we do this under only one seed. When later modules are run under different seeds but the earlier module (say ASSETS) is run under only the one seed, then the between-household variability introduced by that one seed appears to be

Table of Contents

part of sampling variability. When we later run that module under different seeds, this variability is correctly attributed to stochastic simulation and contributes to the estimate of σ_s^2.

This points out the importance of doing multiple runs of all parts of a microsimulation model in which stochastic variability is used, if the relative importance of each error source is desired. We note that in the case of the MATH-CPS model, there are earlier components of the model which used stochastic simulation but which we cannot access, such as imputing missing data. Thus estimates of σ_s^2 will be underestimates of the true stochastic simulation uncertainty, and estimates of σ_h^2 will in general overestimate of the true sampling variability.

All subsequent experiments are run using all 10 MATH-CPS model modules.

8.2 Estimates of sampling and stochastic uncertainty

As pointed out in Section 7.1, we used a nested design in part because we expected stochastic simulation uncertainty to be small relative to sampling variability, and under this assumption a nested design gives an adequate estimate for σ_h^2. Since this assumption turned out to be false when running all 10 modules, we can use a crossed design to help get a better estimate of σ_h^2.

In the nested design the estimator of σ_h^2 is

$$\hat{\sigma}_h^2 = \frac{1}{S}\left(\frac{n(J-1)^2}{J}MS(JK) - (J-1)\hat{\sigma}_s^2\right) \tag{1}$$

In the crossed design the estimator of σ_h^2 is

$$\hat{\sigma}_h^2 = \frac{1}{S}\left(\frac{n(J-1)^2}{J}MS(JK) - \hat{\sigma}_s^2\right) \tag{2}$$

The crossed design estimates σ_h^2 with smaller variability than the nested design, for the following reason. Since σ_s^2 has the same expectation in (1) and (2), the second term in (1) has larger expectation. Also σ_h^2 has the same expectation in both designs, so the first term in (1) must also have a larger expectation than in (2). Since MS(JK) has $(J-1)$ degrees of freedom in both designs, but the estimator of σ_h^2 in (1) involves subtracting two pieces

Table of Contents

each larger in expected magnitude than the two pieces in (2), the estimator for σ_h^2 in the nested design has larger variability than in the crossed design.

Although the crossed design is to be preferred for estimating σ_h^2, the crossed design doesn't estimate σ_s^2 as well as the nested design. The crossed design uses only S seeds whereas the nested design uses SJ seeds. In the crossed design, each household is simulated under S seeds, whereas in the nested design each household is simulated under $S(J-1)$ seeds.

In the ANOVA table corresponding to a crossed design, there are sums of squares for jackknife, seed, and jackknife crossed with seed. Since we use the same model throughout, when running this 2-factor experiment we only have variance components for sampling and seed. Thus the crossed design gives two estimators for σ_s^2, one from MS(seed) and the other from MS(JK $\times$ seed). In both designs, the estimate of σ_h^2 involves subtracting the estimate of σ_s^2 (see Table 29 and Table 32). Thus the crossed design also gives two estimators for σ_h^2 since there are two estimates of σ_s^2 which could be used.

We ran several experiments using the nested design, as well as several experiments with the crossed designs. The former are summarized in Table 9, and the latter in Table 10. For the nested design, we used a jackknife leaving out every 20th record, and in other experiments leaving out every 19th record. We also ran experiments in which we jackknifed rotation group. For each

Table of Contents

Table 9: Summary results from experiments using stochastic and sampling variability, from a nested design. Note that σ_h^2 is proportional to MS(JK) minus MS (seed).

JK group	random				rotation	
number of JK groups	20	20	19	19	8	8
number of seeds per jackknife	4	4	4	4	10	10
seed set	A	B	A	B	A	B
$\hat{\beta}$	9.215	9.118	9.256	9.080	9.273	9.131
MS(JK) MS(seed)	.404 .530	.624 .626	.399 .541	.667 .594	.432 .584	.717 .616
$\sqrt{\hat{\sigma}_h^2/n}$ $\sqrt{\hat{\sigma}_s^2/n}$	0 .709	0 .771	0 .716	.561 .750	0 .715	.249 .734
SE($\beta \mid \hat{\beta}$) single run: full run:	 .709 .081	 .771 .088	 .716 .084	 .936 .568	 .715 .085	 .775 .264
% var due to seed single run: full run:	 100 100	 100 100	 100 100	 64.15 2.43	 100 100	 89.7 11.07
Reference: Table number	44	45	46	47	50	51

Table of Contents

Table 10: Summary results from experiments using stochastic and sampling variability, from a crossed design. (1) refers to estimates using MS(seed). (2) refers to estimates using MS(JK $\times$ seed)

JK group	random		rotation	
number of JK groups	20	20	8	8
number of seeds	4	4	10	10
seed set	A	B	A	B
$\hat{\beta}$	9.609	8.876	9.507	8.942
$\sqrt{\sigma_h^2/n}$ (1)	.525	0	.167	.118
$\sqrt{\sigma_h^2/n}$ (2)	.455	0.288	.164	.165
$\sqrt{\sigma_s^2/n}$ (1)	.423	1.026	.646	.841
$\sqrt{\sigma_s^2/n}$ (2)	.674	.713	.655	.758
SE($\beta \mid \hat{\beta}$)				
single run (1):	.674	1.026	.667	.849
full run (1):	.527	.118	.184	.155
single run (2):	.813	.769	.675	.776
full run (2):	.461	.299	.181	.188
% var due to seed				
single run (1):	39.34	100	93.73	98.05
full run (1):	0.85	100	17.59	41.85
single run (2):	68.68	85.98	94.13	95.46
full run (2):	2.81	7.47	18.63	23.10
Reference: Table number:	48	49	52	53

Table of Contents

Table 11: Summary results from experiments using stochastic and simulation variability, from a combination of a nested and a crossed design. The nested component is used to estimate σ_s^2, while the crossed component is used for MS(JK).

JK group	random				rotation			
Number of JK groups	20	20	20	20	8	8	8	8
Number of seeds	4	4	4	4	10	10	10	10
Seed set for MS(JK)	A	A	B	B	A	A	B	B
Seed set for σ_s^2/n	A	B	A	B	A	B	A	B
$\sqrt{\hat{\sigma}_s^2/n}$ from nested design	.709	.771	.709	.771	.715	.734	.715	.734
$\sqrt{\hat{\sigma}_h^2/n}$ from combination	.441	.414	.290	.248	.136	.125	.183	.176
% var due to seed in single run:	72.10	77.62	85.67	90.62	96.51	97.18	93.85	94.56
full run:	3.29	4.36	7.29	11.28	28.31	33.00	17.90	19.90
SE($\beta \mid \hat{\beta}$) single run:	.835	.875	.766	.810	.728	.745	.738	.755
full run:	.448	.423	.301	.263	.161	.153	.202	.197

rather than random group.

We can use a combination of the nested and crossed designs to estimate σ_h^2 with better precision. We can calculate MS(JK) from the crossed design, then using (2), subtract the estimate of σ_s^2 based on the nested design. We did this using the experiments described above. For each of the four experiments using a crossed design, we subtracted σ_s^2 from a nested design (Table 11). For each experiment we first subtracted the corresponding estimate of σ_s^2 using seed set A, then using seed set B.

The estimates of $\sqrt{\sigma_h^2/n}$ were much less variable using this combination design than they had been under either the crossed design alone, or the nested design alone. Since σ_s^2 differs very little from run to run compared to MS(JK), the estimates of σ_h^2 were fairly sensitive to the seed set used for estimating MS(JK), but not very sensitive to the seed set used for $\hat{\sigma}_s^2$.

Table of Contents

Table 11 shows that the estimated sampling error from jackknifing rotation groups is smaller than when jackknifing random groups. This indicates that rotation groups are much more like each other, and each is more representative of the population than are random groups drawn from the CPS sample.

Table 11 also indicates that if the MATH-CPS model were to be run only once (and estimates of σ_h^2 and σ_s^2 were obtained from another source), stochastic error is a more important contributor to posterior uncertainty than is sampling error. When the MATH-CPS model is run several times, as we have done here, the relative importance of stochastic error diminishes.

The fact that stochastic variability is so large relative to sampling variability when running all 10 modules of the MATH-CPS module in a single run, gives hope for producing more precise model estimates. Variability due to stochastic simulation can be reduced by using more seeds and multiple runs (a relatively easy process), whereas to reduce sampling variability it would be necessary to increase the sample size.

Uncertainty due to sampling could be reduced by combining samples from more than one year, effectively increasing the sample size by a factor of two or more. No sample will be exactly representative of a future year for which simulation results are desired, and a sample which consists of a mixture of years would not be representative of any single year. Differential weighting of different years could be used if it was believed that one year was more representative of what the future year of interest will be. For example if the unemployment rate in one sample year was high but was low in another sample year and it was believed that the unemployment rate would rise, one might upweight the records from the sampled year with high unemployment relative to the sampled year with low unemployment. Another approach would be to upweight individual households based on particular characteristics, as we did with households which receive AFDC. In this case having samples from multiple years is advantageous since there is a richer mix of households in a multiple year sample than there is from a single year sample.

For particular estimands and reforms in which sampling variability is an important contributor to posterior variance, combining samples from multiple years might be worth considering. If on the other hand sampling variability is not an important contributor to posterior variance, the cost of processing twice as many records and making different aging adjustments to the different years may outweigh the benefits of combining samples from different years.

Using multiple year samples could be problematic if eligibility rules for the FSP, AFDC, etc. or other provisions differed in the different years. In this

Table of Contents

case one would have to simulate participation in these programs under the baselaw. This would increase the stochastic error compared to using a single sample year.

The relative importance of stochastic variability in the MATH-CPS model also sheds new light on our results for the QC and MATH-SIPP models (Zaslavsky and Thurston, 1994, and Zaslavsky and Thurston, 1995). In the MATH-SIPP model we found sampling variability to be quite important. It is possible that some of what we labeled as sampling variability in these models was actually stochastic variability at the database creation stage or at an earlier part of the model development. More widespread use of multiple imputation for missing data would help to resolve this matter, and would be feasible if built into the original creation of the survey datasets used by microsimulation models.

8.3 Estimates of calibration uncertainty

We analyzed outputs of 5 MATH-CPS model experiments involving calibration. In the first experiment, we divided the AFDC-eligible families into two groups: low benefit families and high benefit families. We will refer to this as the 2-cell calibration scenario. We based the calibration on results from running the MATH-CPS model using all households and the MPR seed. For each possible cutoff between low and high benefit families (separately for AFDC-Basic and AFDC-UP), we calculated weighting factors which would apply to the low and high benefit groups. Most cutoffs required down-weighting the low benefit families (Figure 1 and Figure 2 in Appendix 1). However we decided to pick a cutoff in which no AFDC family was down-weighted. For the AFDC-Basic families, this meant the high benefit families were up-weighted by 3.801, whereas the high benefit AFDC-UP families were up-weighted by 11.026. In both cases, the low benefit families had a weighting factor of about 1 (Table 12). In order to keep the sum of total household weights on the file constant, non-AFDC households were down-weighted accordingly.

Eligibility and participation for AFDC is determined at the family level. Sampling and weighting apply to households. This was problematic for households which contained more than one family because not all families within a single household received the same type of AFDC. In some households, one family received AFDC and another did not. Consequently we did not exactly meet the control totals for all four targets, but were off by at most 5%.

In the first experiment these weighting factors were applied to MATH-CPS

Table of Contents

Table 12: Details of 2 cell calibration method, based on one seed and the full sample

	benefit class	benefit cutoffs	number of households	weighting factors
AFDC-Basic households	low	11-820	2346	1.0013
	high	821-1625	92	3.8057
AFDC-UP households	low	10-900	137	1.0032
	high	901-1057	6	11.0258
Non-AFDC	all households		56390	.9956

model runs in which we jackknifed random groups and used different seeds. Since eligibility for AFDC depends on a particular seed and the total number of eligible families depends on the particular jackknife group, the weighting factors we developed based on the full sample and the MPR seed meant that the AFDC control totals were not exactly met in any individual run. Also the total sum of weights on the file varied from run to run.

We ran the MATH-CPS model 40 times with the standard MPR calibration, and 40 times with our alternate calibration method as described above. Each of the 40 runs used 10 jackknife samples with 4 seeds nested. The estimated calibration effect from this design was -.329. A negative calibration effect means that the estimated percent change in benefits is smaller when using our alternate method of calibration as compared to the standard MPR calibration. Calibration accounted for 47.07% of the total variability (Tables 54 and 13).

The disadvantage of the particular 2-cell calibration scenario we used is that the up-weighting effectively only applies to a few households. Only 6 households which receive AFDC-UP (4%), and 92 households which receive AFDC-Basic (4%) were up-weighted. Consequently we decided to look at the class of calibrations involving 3 groups of households in each of the two AFDC categories. We will call these 3-cell calibration scenarios. A summary of all 5 calibration experiments can be found in Table 13.

In the first of the 3-cell calibration scenarios, we based the calculation of the weighting factors on results from running the MATH-CPS model using all households and the MPR seed, as we had for the 2-cell calibration scenario. We decided to down-weight low benefit households in both groups. The 148 lowest benefit AFDC-Basic households (comprising 9% of AFDC-Basic households) were given a weighting factor of .75, while the 6 lowest benefit AFDC-UP households (4% of the AFDC-UP households) were given a weighting factor of .50. Looking at all possible cutoffs between the medium

Table of Contents

Table 13: Summary results from calibration experiments. Each experiment used 10 MATH-CPS model modules and was run under reform 1.

Jackknife group	random				rotation
Number of resamples	10	10	10	10	8
Number of seeds	4	4	4	4	4
Number of benefit groups	2	3	3	3	3
Weighting factors developed from	one run	one run	each run	median of 80 runs	each run
$\hat{\beta}$	8.954	8.816	8.475	8.466	8.420
$\hat{\beta}_C$	-.329	-.488	-.829	-.839	-.900
$\mathrm{SE}(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$					
single run:	0.768	0.801	1.035	1.044	1.116
full run:	0.480	0.502	0.842	0.848	0.953
% var due to calibration (main effect)					
single run:	18.40	37.11	64.24	64.58	65.09
full run:	47.07	94.51	97.01	97.93	89.22
Table number	54	55	56	57	58

Table of Contents

Table 14: Details of 3 cell calibration method, based on one seed and the full sample

	benefit class	benefit cutoffs	number of households	weighting factors
AFDC-Basic households	low	11-100	148	0.7500
	medium	101-442	1419	1.0003
	high	443-1625	871	1.4374
AFDC-UP households	low	10-100	6	0.5000
	medium	101-600	94	0.7490
	high	601-900	44	2.7872
Non-AFDC	all households		56389	0.9938

and high benefit households in both AFDC groups, we decided on one which used weighting factors of 1.00 for the medium benefit AFDC-Basic households, and .75 for the medium benefit AFDC-UP households. This meant that the 871 high benefit AFDC-Basic households (36%) had a weighting factor of 1.44, and the 44 high benefit AFDC-UP households (31%) had a weighting factor of 2.79 (Table 14). Compared to the 2-cell calibration scenario, a much larger number of high benefit households in each AFDC group were up-weighted, and by a smaller amount. This means that results of this 3-cell calibration scenario do not rely so heavily on a very few households.

Results from this experiment gave an estimated calibration effect of -.488 for the full run. Calibration accounted for 94.51% of the variability for the full run (Tables 13 and 55).

We also checked to see what effect these methods of calibration had on the other target quantities. As can be seen from Table 15, calibration using either the 2 cell method or the 3 cell method generally brought other totals or percentages of interest closer to the targets. In particular, the percentage of FSP households with AFDC was very close to the target when our method of calibration was used, as was the percentage of FSP households with children aged 5-17. This is to be expected, since up-weighting AFDC households relative to non-AFDC households should increase the proportion of AFDC households in the FSP, and since AFDC families have children, this also increases the number of households with children in the FSP. It is encouraging to note that these numbers were quite close to the target numbers.

In our third calibration experiment, we decided to re-calibrate each run of the MATH-CPS model. This meant the weighting factors were recalculated for each jackknife sample and seed under which we ran the model, so that the 4 AFDC control totals would be met for that particular run (as much as

Table of Contents

Table 15: Comparison of baselaw targets and microsimulation estimates using two different calibration methods

Quantity	Target	standard calibration	our calibration with 2 groups	3 groups
Weighted number of units participating in FSP (x1000)	10,864	9,204	9,532	9,657
Weighted total benefits for the FSP (x1000)	2,029,742	1,679,354	1,792,352	1,812,331
Percentage of FSP households with AFDC	40	37.0	39.4	40.3
Percentage of FSP households with SSI	19	23.2	22.7	22.5
Percentage of FSP households with GA	8	6.0	5.9	5.9
Percentage of FSP households with earners	20	23.1	23.0	22.3
Percentage of FSP households with elderly	15	21.7	21.2	20.7
Percentage of FSP households with children (5-17)	43	41.5	43.5	43.6

possible, considering the problem of multiple families within a household). We fixed the weighting factors for the low benefit groups at .75 and .50 for AFDC-Basic and AFDC-UP households respectively, as they had been in the previous experiment. Weighting factors for the medium and high benefit groups were calculated to meet the AFDC control totals. Our expectation was that this would reduce total variability since we would have eliminated variability between runs due to different AFDC control totals.

This experiment also involved 80 runs of the MATH-CPS model, 40 under the standard MPR calibration method, and 40 under our calibration method. Again, the 40 runs were based on 10 jackknife samples and 4 seeds nested in each. Results from calibrating each run apparently indicated the opposite of what we expected. For the full run, the estimated calibration effect was -.829, with calibration accounting for 97.01% of the variability (Tables 13 and 56). Estimates of the calibration effect for individual runs are positively correlated with the weighting factors for the AFDC-Basic groups.

A closer look at the individual weighting factors for each run shows that in general the weighting factors are larger than they were in the previous experiment. For AFDC-Basic households, the weighting factors calculated on the basis of the one MPR seed and all households were lower than any

Table of Contents

Table 16: Details of 3 cell calibration method when calibration is done separately for each run

	benefit class	benefit cutoffs	number of households	weighting factors
AFDC-Basic households	low	11-100	119 to 154	.750
	medium	101-442	1209 to 1311	1.085 to 1.197
	high	443-1625	753 to 813	1.484 to 1.659
AFDC-UP households	low	10-100	1 to 10	0.500
	medium	101-600	60 to 99	0.712 to 1.596
	high	601-900	23 to 41	2.659 to 5.149
Non-AFDC	all households			0.988 to .990

of the weighting factors calculated in 40 runs using jackknife samples and different seeds. For AFDC-UP families, the weighting factors based on the MPR seed and full sample were near the low end of the weighting factors calculated from the 40 runs (Table 16).

Since the 40 calibrated runs of this latest experiment were based on the same set of 10 jackknife samples and 4 seeds nested in each as the 3 cell calibration scenario in which calibration was done only once, there is a natural pairing of runs. We expected that individual estimates of the calibration effect would be positively correlated, and we found this to be the case (Figure 3 in Appendix 1). Figure 3 in Appendix 1 also shows that the estimated calibration effect based on the MPR seed and the full sample was smaller than any individual calibration effects for the 40 runs.

These results suggested a fourth experiment involving calibration. We wanted to compare results of calibrating each run with an experiment in which the weighting factors were only calculated once, but in which the weighting factors were near the center of the distributions of weighting factors generated from calibrating each run. The fourth experiment used the median weighting factors from the experiment in which each run was calibrated separately (Table 17).

Results for the fourth experiment were extremely similar to the third experiment, in which calibration was done for each run. For the full run, the estimated calibration effect was -.839 (as compared to -.829), with calibration accounting for 97.93% of the total variability (as compared to 97.01%) (Tables 13 and 57). A plot of estimated calibration effects for each run when calibration is done once versus when calibration is done each time (Figure 4) shows a positive correlation, with about equally many points above and many points below the line of equality.

Table of Contents

Table 17: Comparison of weighting factors in 3 cell calibrations based on the MPR seed versus the median of 40 runs

	benefit class	weighting factors from one run	weighting factors from median of 80 runs
AFDC-Basic households	low	.750	.750
	medium	1.000	1.142
	high	1.437	1.566
AFDC-UP households	low	.500	.500
	medium	.749	1.102
	high	2.787	3.819
Non-AFDC	all households	.994	.989

The final experiment involving calibration used a jackknife of rotation groups instead of random groups, and involved calibrating each run. Results of this experiment are quite similar to results of the analogous experiment using a jackknife of random groups. The estimated calibration effect is -.900 (as compared to -.829), with calibration accounting for 89.22% of the posterior variance (Tables 13 and 58).

In conclusion, simulation results using the MPR seed are outlying with respect to the estimate of the calibration effect. Otherwise, the estimates from other seeds are quite consistent.

8.4 Estimates of five sources of uncertainty

We ran the MATH-CPS model incorporating all five sources of uncertainty to determine the absolute and relative sizes of each source of uncertainty. In all cases we used a jackknife of rotation group to estimate sampling variability. To simplify the experiments we used two seeds nested in each rotation group to estimate stochastic variability, although the results from Section 8.2 suggest that a combination of the crossed and nested design could improve estimates of sampling effects, σ_h^2, σ_{hc}^2, etc.

We used the calibration method based on dividing each of the two types of AFDC (AFDC-Basic and AFDC-UP) into three groups. We did not calibrate each run separately, since results from calibration experiments indicated that calibrating each run separately gave results very similar to runs which were not calibrated separately.

We used two unemployment rates, 6.4%, and 4.3%, as described earlier. We used two methods of allocating earned income to months, also as described earlier.

8.4.1 Results from reform 1, using 2 seed sets

We did 12 complete (five-factor) MATH-CPS model experiments. The first two were done using the original reform (reform 1). The first run used seed set A, and the second run used seed set B. The other 10 MATH-CPS model runs were done using different reforms.

As in experiments involving only stochastic and sampling variability (Section 8.2), σ_h^2 was not well estimated in the full MATH-CPS run (Table 18, and Tables 59 and 60). In the first reform using seed set A, $\sqrt{\hat{\sigma}_h^2/n}$ was .674, whereas using seed set B, $\sqrt{\hat{\sigma}_h^2/n}$ was 0. Using seed set A, $\sqrt{\hat{\sigma}_s^2/n}$ was .530, which is somewhat smaller than estimates from other experiments, whereas using seed set B, $\sqrt{\hat{\sigma}_s^2/n}$ was .787, about the same as previous estimates.

Due to the very different estimates of σ_h^2 in the two runs of the first reform, the estimated relative importance of sampling and stochastic error to posterior variance were quite different. Sampling accounted for 35% of the posterior variance from a full experiment (29% from a single run) using seed A, but 0% using seed B. Stochastic uncertainty accounted for 1.5% of the posterior variance using seed A (18% from a single run), but 6% under seed B (48% from a single run). Although the estimated calibration effect was slightly smaller using seed set B, calibration accounted for a greater percent of the posterior variance using seed B (92% for a full run) than using seed A (51% for a full run). For a single run, these numbers reduced to 43% for seed A and 49% for seed B.

Calibration was the only fixed effect that accounted for more than 2% of the posterior variability for this reform. Uncertainty about the unemployment rate and about method of income allocation were not important contributors to total posterior variance.

We compared the original reform (reform 1) with 10 other reforms (summarized in Table 19). These reforms are grouped into two sets: reforms 1-6 and reforms 7-11. Reforms 1-6 were run first, and results from these reforms suggested that another set of reforms be run which might have more uncertainty. We present results from these two sets of reforms separately. The reforms were run using only seed set A.

Table of Contents

Table 18: Summary results from the full 5-factor model under reform 1, using two different seed sets.

	seed set A	seed set B
$\sqrt{\hat{\sigma}_h^2/n}$	.674	0.0
$\sqrt{\hat{\sigma}_s^2/n}$	.530	0.787
$\hat{\beta}$	8.468	8.203
$\hat{\beta}_C$	-.821	-.796
posterior SE:		
single run	1.259	1.136
full run	1.145	0.832
% var due to sampling (main effect):		
single run	28.660	0.0
full run	34.638	0.0
% var due to seed (main effect):		
single run	17.717	47.920
full run	1.529	6.392
% var due to calib (main effect):		
single run	42.555	49.120
full run	51.431	91.732

8.4.2 Results from reforms 1-6

In reform 2, the earnings deduction is 50%, as compared to 20% in the baselaw. We expected that for this reform, the unemployment rate and possibly the method of income allocation should be a more important contributor to posterior variance than under the original reform.

In reform 3, the shelter deduction cap is eliminated. This reform and others involving changes in shelter deduction are frequently analyzed at MPR.

Reforms 4 and 5 both affect the treatment of assets. Since both vehicular and financial assets are imputed in the MATH-CPS model, we expected stochastic uncertainty to be important in these reforms. In reform 4, the first vehicle is not counted as an asset, so people cannot become ineligible for the food stamp program due to the value of their first car. Under the baselaw, any amount above $4600 in the value of the first car is counted as an asset.

Reform 5 is an extreme asset reform in which asset limits are eliminated completely. This means that households which are eligible for the food stamp program based on other criteria, are still eligible despite having any amount of assets. This reform has been run at MPR many times, not as a realistic option, but to determine how many people are kept off the food stamp program because of their assets.

In reform 6, the earnings deduction is increased to 50% (as compared to 20% in the baselaw), but only for FSP households which participate in AFDC. This reform is similar to reform 2 in which the 50% earnings deduction applies to all households. Reform 6 is an example of a type of reform that may be requested under the new welfare block grants, in which the FSP benefits for families applying for TANF (Temporary Assistance to Needy Families)

Table of Contents

Table 19: Description of reforms under which the full MATH-CPS model was run

Reform number	Description
Baselaw	(used as reference for all reforms) Asset limit for elderly households = $3000 Asset limit for non-elderly households = $2000 Dependent child care cap for 48-states: for children under 2 years = $200 for children 2 years and over = $175 Shelter cap multiplier = 1 Earnings deduction = 20% Value of first vehicle over $4600 counted as asset
1	Asset limit for elderly households = $5000 Asset limit for non-elderly households = $5000 No dependent child care cap for 48-states Shelter cap multiplier = 0.5
2	Earnings deduction = 50%
3	No shelter deduction cap
4	Value of first vehicle not counted as asset
5	No asset limit
6	Earnings deduction = 50% for AFDC households only
7	Asset limit for elderly households = $2150 Asset limit for non-elderly households = $2150
8	Shelter deduction = shelter expenses over 35% of gross income
9	Earnings deduction = 75% for AFDC-UP households only
10	Earnings deductions = 34% for households on FSP for 6 months or less No earnings deductions for households on FSP for more than 6 months
11	Value of first vehicle over $10,000 counted as asset Asset limit = $1075 for non-elderly households on FSP for more than 6 months

Table of Contents

Table 20: Estimates of main, 2-way and 3-way effects in MATH-CPS model reforms 1-6. All are run jackknifing rotation group, with 2 seeds nested. An asterisk indicates the effect is more then 2 SE's above 0 in the full run.

	reform number					
	1	2	3	4	5	6
$\hat{\beta}$	8.468 *	10.435 *	2.838 *	5.314 *	26.056 *	3.081 *
$\hat{\beta}_C$	-.821 *	-.502 *	.086 *	-.355 *	-1.705 *	.070 *
$\hat{\beta}_E$	-.081	.803 *	-.049	-.047	-.123	.279 *
$\hat{\beta}_I$	.037	-.425 *	-.040 *	-.019	.018	-.039 *
$\hat{\beta}_{CE}$	.017 *	-.026 *	-.008	.007 *	.016	.007 *
$\hat{\beta}_{CI}$	-.003	.027 *	.003	.001	.001	-.004
$\hat{\beta}_{EI}$	.004	-.033	-.002	.002	.002	.001
$\hat{\beta}_{CEI}$	.000	.003 *	.000	.000	.001	.000

Table 21: Estimates of simulation and sampling variability in MATH-CPS model reforms 1-6. All are run jackknifing rotation group, with 2 seeds nested, using seed set A.

	reform number					
	1	2	3	4	5	6
$\hat{\beta}$	8.468	10.435	2.838	5.314	26.056	3.081
$\sqrt{\hat{\sigma}_h^2/n}$	0.674	0.0	0.044	0.0	1.774	0.0
$\sqrt{\hat{\sigma}_s^2/n}$	0.530	0.349	0.124	0.587	0.659	0.232
Posterior SE:						
single run	1.259	1.110	0.193	0.745	2.633	0.385
full run	1.145	1.051	0.145	0.476	2.548	0.300
Coefficient of variation						
single run:	.149	.106	.068	.140	.101	.125
full run:	.135	.101	.051	.090	.098	.097
% var due to sampling (main effect):						
single run:	28.660	0.0	5.512	0.0	45.403	0.0
full run:	34.638	0.0	9.133	0.0	48.481	0.0
% var due to seed (main effect):						
single run	17.717	9.898	41.005	62.195	6.259	36.204
full run	1.529	0.789	5.192	10.857	0.477	4.265

Table of Contents

Table 22: Percentage of posterior variance in a full run due to major sources of error in MATH-CPS model reforms 1-6. All are run jackknifing rotation group, with 2 seeds nested, using seed set A.

	reform number					
	1	2	3	4	5	6
sampling (J)	34.638	0.000	9.133	0.000	48.481	0.000
seed (S)	1.529	0.789	5.192	10.857	0.477	4.265
calibration (C)	51.431	22.766	35.247	55.545	44.777	5.492
unemployment (E)	0.497	58.354	11.614	0.956	0.233	86.843
income (I)	0.103	16.337	7.683	0.153	0.005	1.696
C × E	0.022	0.062	0.272	0.021	0.004	0.055
C × I	0.001	0.065	0.037	0.000	0.000	0.016
E × I	0.001	0.099	0.023	0.001	0.000	0.002
C× E × I	0.000	0.001	0.000	0.000	0.000	0.000
E × J	9.671	0.000	28.994	22.516	4.335	0.000
I × J	1.191	1.004	0.000	6.626	0.915	0.000

reform to reform, as expected (Table 20). Reform 5, which eliminates asset limits, resulted in an estimated 26% percent change in benefits. Reform 3, in which the shelter deduction cap is eliminated, gave less than a 3% change in benefits. In all cases the estimated percent change in benefits was positive, indicating a larger simulated total benefit under the reform than under the baselaw.

Calibration was the only fixed effect which was significant for all reforms, by which we mean that $\hat{\beta}_C$ is more than 2 standard errors away from 0. For all reforms except for the two reforms involving earnings deductions (reforms 2 and 6), the calibration effect was the only significant and large effect, and except for reform 6, it was estimated to be larger than any other fixed effect. For most reforms, $\hat{\beta}_C$ was negative indicating that the percent change in benefits is smaller when using our method of calibrating AFDC totals as compared to the standard MPR method of AFDC calibration. This means that for the reforms with a negative $\hat{\beta}_C$, the total reform benefits are closer to the total baselaw benefits when AFDC families are up-weighted. For reform 6, in which a 50% earnings deduction only applies to AFDC families, the difference in benefits is larger when AFDC families are up-weighted.

Only in reforms 2 and 6, in which the earnings deduction is 50% instead of 20% under the baselaw (for AFDC families only, in reform 6), were uncertainty about the unemployment rate and uncertainty about method of income allocation important contributors to posterior variance. As expected, a reform which increases the earnings deduction is sensitive to what the pro-

Table of Contents

jected unemployment rate is, and to the distribution of earned income over the year. Under reform 2, $\hat{\beta}_C$ was $-.502$, $\hat{\beta}_E$ was .803, and $\hat{\beta}_I$ was $-.425$, all three of which were statistically significantly different from zero. In reform 6, $\hat{\beta}_C$ was .070, $\hat{\beta}_E$ was .279, and $\hat{\beta}_I$ was $-.039$.

As can be seen from Table 21, σ_h^2 is not well estimated from the nested design, nor is the percentage of posterior variance due to sampling variability. The estimated magnitude of stochastic standard error in a single run ($\sqrt{\sigma_s^2/n}$) was quite large for the two asset reforms (reform 4 and reform 5). The original reform (reform 1) also had large stochastic uncertainty relative to reforms 2, 3 and 6, which is not surprising since the original reform also involves an increased asset limit, while reforms 2, 3, and 6 do not. In reform 5 in which σ_s^2/n is large, $\hat{\beta}$ is also very large, and the relative importance of stochastic uncertainty is quite small. In contrast, for reform 4, in which vehicles are not counted as assets, stochastic uncertainty accounts for over half the total uncertainty in a single run, and over 10% of the uncertainty in a full run. For this reform in particular, multiple runs of the model using different seeds are especially worthwhile. Running the model under multiple seeds decreased the posterior standard error by one third.

The estimated posterior standard error varied quite a bit between the different reforms. For a full run, the posterior SE ranged from 0.145 for reform 3 to 2.548 for reform 5. As a percentage of the point estimate, however, the standard error was remarkably similar across reforms. The coefficient of variation (the ratio of the posterior standard error to $\hat{\beta}$) ranged from 5.1% to 13.5% in the full run (6.8% to 14.9% in a single run).

In Table 22, which compares the relative importance of major sources of uncertainty, we can see that calibration method is an important contributor to posterior uncertainty for all reforms except for reform 6. For reforms 1 through 5, calibration accounted for from 23% to 56% of the total variance when using a full design (only 5.5% for reform 6). Using a full design, unemployment accounted for most (87%) of the uncertainty under reform 6, more than half (58%) of the uncertainty under reform 2, 12% of the uncertainty under reform 3, and was negligible under the other reforms. Uncertainty about income allocation method was important for reforms 2 and 3, but not for the other reforms.

Surprisingly, the interaction of employment with jackknife was an important contributor to posterior variance for several reforms, accounting for up to 29% of the total posterior variance. This means that with the available data it is hard to estimate the effect of a changed unemployment rate on the percent change in FSP benefits, but that the possibility exists that the effect

Table of Contents

is quite large.

8.4.3 Results from reforms 7-11

Like each reform in the first set, each reform in the second set is either an actual reform analyzed by FCS or a hypothetical reform with provisions similar to those typically analyzed by FCS. However, the reforms in the second set were selected because we thought the uncertainty in the estimates from these reforms would be substantially larger than from the first set of reforms.

In reform 7 the asset limit for all households is $2150 which is slightly more than the present $2000 for non-elderly households, and less than the present $3000 for elderly households. We expected that stochastic simulation uncertainty would be large in this reform since assets are imputed.

Reform 8, like reform 3, incorporates a change in shelter deduction. In reform 8, the shelter deduction is the shelter expenses over 35% of gross income. We expected that this reform would have larger uncertainty than reform 3 since it relies on gross income, which in turn depends on employment status.

Reforms 9 and 10, like reforms 2 and 6, involve earnings deductions. In reform 9, the earnings deduction is 75%, but just for AFDC-UP households. In reform 10, the earnings deduction is 34%, but just for households on the FSP for 6 months or less, and the earnings deduction is 0% for households on the FSP for more than 6 months. With all earnings deduction reforms, we expect error due to uncertainty about the unemployment rate to be relatively large. Since relatively few households are on AFDC-UP, we expect sampling uncertainty to be relatively large in reform 9.

In reform 11 the first $10,000 of the value of the vehicle is not considered an asset, and the asset limit for non-elderly households on the FSP for more than six months is reduced to $1075. We expect stochastic simulation uncertainty to be large in this reform due to imputation of financial and vehicular assets.

Detailed results from running the second set of reforms are shown in Tables 66, 67, 68, 69, and 70. Summaries are given in Tables 23, 24, and 25. Table 23 shows estimates of the fixed effects, Table 24 compares sampling and stochastic error, and Table 25 gives the relative contributions to posterior variance from important sources of uncertainty.

The percent change in benefits under the second set of reforms was estimated much less precisely than under the first set. The point estimates for percent

Table of Contents

Table 23: Estimates of main, 2-way and 3-way effects in MATH-CPS model reforms 7-11. All are run jackknifing rotation group, with 2 seeds nested. An asterisk indicates the effect is more then 2 SE's above 0 in the full run.

	reform number				
	7	8	9	10	11
$\hat{\beta}$	0.308 *	0.215	0.302	0.166	0.291 *
$\hat{\beta}_C$	-.028	.077 *	.056 *	-.057 *	-.047
$\hat{\beta}_E$	-.022	-.058 *	.000	-.154 *	-.076
$\hat{\beta}_I$	-.002	.010	-.003	-.030	.004
$\hat{\beta}_{CE}$	.002	.002	.002	.004	.005
$\hat{\beta}_{CI}$	.000	-.001	-.003	.006 *	-.001
$\hat{\beta}_{EI}$	.004	.000	-.001	.009	.004
$\hat{\beta}_{CEI}$	.000	.000	-.001	.001	.000

Table 24: Estimates of simulation and sampling variability in MATH-CPS model reforms 7-11. All are run jackknifing rotation group, with 2 seeds nested, using seed set A.

	reform number				
	7	8	9	10	11
$\hat{\beta}$	0.308	0.215	0.302	0.166	0.291
$\sqrt{\hat{\sigma}_h^2/n}$	0.0	0.130	0.163	0.182	0.0
$\sqrt{\hat{\sigma}_s^2/n}$	0.299	0.045	0.040	0.159	0.422
Posterior SE:					
single run	.323	.172	.189	.315	.563
full run	.114	.165	.174	.266	.378
Coefficient of variation					
single run:	1.049	.800	.626	1.898	1.935
full run:	.370	.767	.576	1.602	1.299
% var due to sampling (main effect):					
single run:	0.0	57.016	74.448	33.273	0.0
full run:	0.0	62.003	87.303	46.520	0.0
% var due to seed (main effect):					
single run	85.709	6.910	4.519	25.415	56.145
full run	49.626	0.537	0.379	2.538	8.888

Table 25: Percentage of posterior variance in a full run due to major sources of error in MATH-CPS model reforms 7-11. All are run jackknifing rotation group, with 2 seeds nested, using seed set A.

	reform number				
	7	8	9	10	11
sampling (J)	0.000	62.003	87.303	46.520	0.000
seed (S)	49.626	0.537	0.379	2.538	8.888
calibration (C)	6.076	22.059	10.480	4.645	1.549
unemployment (E)	3.848	12.493	0.000	33.500	4.019
income (I)	0.031	0.382	0.034	1.273	0.010
C × E	0.033	0.015	0.012	0.025	0.016
C × I	0.000	0.002	0.029	0.044	0.000
E × I	0.109	0.000	0.001	0.111	0.011
C× E × I	0.001	0.000	0.002	0.000	0.000
E × J	0.000	2.001	0.000	2.821	74.165
I × J	24.080	0.000	0.000	6.772	5.821

change in benefits for reforms 7-11 were all positive and all smaller than those for reforms 1-6. The impacts for reforms 7-11 (less than half a percent, compared to 2.8 to 26 percent in reforms 1-6) are more typical of the impacts estimated by FCS on a routine basis.

As anticipated, the posterior standard errors for reforms 7-11 were much larger relative to the point estimates than was the case with reforms 1-6. The coefficients of variation ranged from .37 to 1.60 for a full run (.63 to 1.94 for a full run), which is substantially larger than for the first set of reforms. In addition, the large standard errors relative to the point estimates means that in all five reforms, in a single run, we cannot be very certain whether the percent change in benefits is positive or negative. The approximate 95% posterior intervals for percent change in benefits included 0 for all five reforms in a single run, and for all but reform 7 in a full run.

Stochastic simulation uncertainty was particularly large for reforms 7 and 11, the two reforms which involve assets. In reform 7, $\sqrt{\hat{\sigma}_s^2/n} = .299$, which was almost equal to $\hat{\beta} = .308$. In reform 11, $\sqrt{\hat{\sigma}_s^2/n} = .422$ which was much larger than $\hat{\beta} = .291$ (Table 24). Thus stochastic variability alone was of about the same magnitude as the point estimate for these two reforms. In a single run, stochastic simulation accounted for 86% of the posterior variance for reform 7, and 56% of the posterior variance for reform 11. Under the full design, these numbers were reduced to 50% and 9% respectively. In reform 11, the major contribution to posterior variance was the employment × jackknife

Table of Contents

interaction, accounting for 74% of the posterior variance.

Due to the extremely large stochastic simulation uncertainty in reforms 7 and 11, it is not surprising that none of the fixed effects ($\hat{\beta}_C, \hat{\beta}_E$, etc.) were significantly different from 0. These were the only two reforms out of all 11 reforms in which there were no significant fixed effect.

In reforms 8, 9, and 10, the calibration effect was more than 2 standard errors from 0. Calibration accounted for 5% of the posterior variance in the full run in reform 10, 10% in reform 9, and 22% in reform 8. The relative importance of calibration for the second set of reforms is substantially smaller than for the first set of reforms (except for reform 6 in which uncertainty about the unemployment rate was very important). This is at least in part due to the large absolute and relative importance of stochastic variability in reforms 7 and 11, and sampling variability in reforms 8, 9 and 10.

In reform 9, sampling accounted for 87% of the posterior variance. This is not surprising since this reform only involves AFDC-UP households, and there are very few such households in the dataset. Sampling accounted for 62% of the posterior variance for reform 8, and 47% for reform 10.

Reforms 8, 9 and 10 involve earnings in some way, and household earnings are affected by the employment status of people in the household. In reform 8, the shelter deduction depends on gross income, whereas reforms 9 and 10 involve earnings deductions. The unemployment effect, $\hat{\beta}_E$ was significantly different from 0 in reforms 8 and 10. In reform 10, $\hat{\beta}_E$ was almost as large in magnitude as $\hat{\beta}$ ($\hat{\beta}_E = -.154$, and $\hat{\beta} = .166$). In no other reform was a fixed effect nearly so close in magnitude to the point estimate.

8.4.4 General comparisons of all 11 reforms

The eleven reforms we considered are summarized briefly in Tables 26 and 27. Of the reforms we considered, four involved asset limits (reforms 1, 5, 9 and 11). In reforms 1 and 5 the asset limit was increased as compared to the baselaw, whereas in reforms 9 and 11 the asset limit was decreased for some subset of households. The increase in stochastic simulation variability, which we expected for asset imputations, was very large only in the reforms in which the asset limit was smaller in the reform than in the baselaw. There may be relatively few households with imputed assets over the baselaw limits which are otherwise eligible for the FSP. If this is so, we expect sampling variability to be large in reforms with an increased asset limit, and in fact this is what we found.

Table of Contents

Table 26: Estimates of the standard error and coefficient of variation in all 11 MATH model reforms. Numbers are for a single run, with full run numbers in parentheses.

Ref	$\hat{\beta}$	CV	% var due to seed (main effect)	reform about
1	8.47	.15 (.14)	17.72 (1.53)	asset limits, other
2	10.44	.11 (.10)	9.90 (0.79)	earnings deduction (all hhs)
3	2.84	.07 (.05)	41.01 (5.19)	shelter deduction
4	5.31	.14 (.09)	62.20 (10.86)	first vehicle
5	26.06	.10 (.10)	6.26 (0.48)	asset limits
6	3.08	.13 (.10)	36.20 (4.27)	earnings deduction (AFDC)
7	0.31	1.05 (.37)	85.71 (49.63)	asset limits
8	0.22	.80 (.77)	6.91 (0.54)	shelter deduction
9	0.30	.63 (.58)	4.52 (0.38)	earnings deduction (AFDC-UP)
10	0.17	1.90 (1.60)	25.42 (2.54)	earnings deduction
11	0.29	1.94 (1.30)	56.15 (8.89)	first vehicle, asset limits

Table 27: Percentage of posterior variance in a full run due to major sources of error in all 11 MATH model reforms.

Ref	J	S	C	E	I	E×J
1	34.64	1.53	51.43	0.50	0.10	9.67
2	0.00	0.79	22.77	58.35	16.34	0.00
3	9.13	5.19	35.25	11.61	7.68	28.99
4	0.00	10.86	55.55	0.96	0.15	22.52
5	48.48	0.48	44.78	0.23	0.01	4.36
6	0.00	4.27	5.49	86.84	1.70	0.00
7	0.00	49.63	6.08	3.85	0.03	0.00
8	62.00	0.54	22.06	12.49	0.38	2.00
9	87.30	0.38	10.48	0.00	0.03	0.00
10	46.52	2.54	4.65	33.50	1.27	2.82
11	0.00	8.89	1.55	4.02	0.01	74.17

Earnings deductions were involved in four reforms: reforms 2, 6, 9 and 10. As noted earlier, with the exception of reform 9 which involved only a few households on AFDC-UP, uncertainty about the unemployment rate was a major contribution to posterior variance for these earnings deduction reforms.

Two reforms, reforms 4 and 11, involve vehicular assets. In both, the ratio of stochastic variability to $\hat{\beta}$ was relatively large. Although stochastic error was relatively large, it was the employment $\times$ jackknife interaction which was a surprisingly important source of uncertainty in these reforms. We hypothesize that there is a small number of people who are unemployed part of the year and have expensive cars, and due to small numbers, these people are not evenly distributed across rotation groups.

One possible explanation for the large employment $\times$ jackknife interactions is that the difference in unemployment rate has a very different effect on the percent change in FSP benefits for different people. When averaging very different effects (possibly also differing in sign) over many people, this would give a small estimated effect, with a large standard error, which is generally what we found.

Another possible explanation is that people who are unemployed part of the year tend to be clustered in certain small areas, thus comprising a relatively large percentage of the sample in some rotation groups but not in other rotation groups.

Whether incorporation of multiple seeds into the MATH-CPS model would be useful depends on what the outputs of interest are, and on what reform is of interest. If the model is run under one seed only, our best estimate of the percentage of posterior variance due to seed, for the estimand of percent change in benefits, ranged from 4.5% (reform 9) to 85.7% (reform 7). For the reform in which seed was apparently most important (reform 7), the posterior standard error decreased from 0.32 to 0.11 when multiple seeds were used. Thus using multiple seeds can give a substantial reduction in the posterior standard error for some reforms. We note that using multiple seeds is beneficial particularly for reforms involving assets, and these reforms could be tested using the MATH-SIPP model in which assets are not imputed.

Using multiple seeds seems most useful when a close examination of one or a few estimands is of interest. However, averaging results over multiple seeds for all entries of tables may produce inconsistent table results. For example, in the gainer/loser table, one estimand is the ratio of the average change in benefits to the average benefit under the base plan. Since both numerator and denominator vary with seed, the average of the ratios will not necessarily agree with the ratio of the averages. We suggest that table entries be initially

Table of Contents

produced from a single seed, but that adding the capacity to average over multiple seeds in the MATH-CPS model would be worthwhile, if feasible. Multiple seeds could then be used for estimands of particular interest, and could be used to determine which estimands and which reforms have large stochastic variability.

9 Conclusions

We now summarize the implications of the results described in Section 8, grouped by the set of experiments to which they apply. In all cases we considered uncertainty in the estimated percent change in benefits under an alternate reform.

In the first set of experiments we incorporated sampling and stochastic error using reform 1, and varied the number of MATH-CPS model modules run. Our conclusions from this set of experiments are:

1. For the main reform we considered (reform 1), the ASSETS module, which imputes financial and vehicular assets, appears to be the primary source of stochastic simulation uncertainty in the MATH-CPS model.
2. If one wants to estimate the relative sizes of sampling and stochastic variability, it is important to run all parts of a microsimulation model which use random numbers, under different sets of random numbers. If parts of the model which use random numbers are not run under different random numbers, the resulting estimate of sampling variability will include variability due to stochastic simulation.

In the second set of experiments, we ran all 10 MATH-CPS modules using reform 1, and investigated more thoroughly the absolute size and relative importance of sampling and stochastic simulation uncertainty. Our conclusions from these experiments are:

1. The estimate of sampling error in the MATH-CPS model cannot be well determined using a design in which seed is nested in jackknife. An estimator for the sampling error based on both a nested and a crossed design gives a more precise estimate of the sampling error than either the nested or crossed design by itself.
2. Based on estimates using a nested and crossed design, stochastic simulation uncertainty is a much more important contributor to total posterior variance than sampling variability when the MATH-CPS model is run only once, for the estimand of percent change in benefits under reform 1.
3. Running the MATH-CPS model many times using different seeds and different jackknife replicates resulted in a sizeable decrease in the uncertainty of the percent change in benefits under reform 1, and would probably reduce the uncertainty in other model outputs as well.

Table of Contents

4. Our estimate of uncertainty due to stochastic simulation is probably too small because we could not model stochastic simulation uncertainty in the creation of the database, such as in imputation of individual missing values in the CPS dataset.

5. The estimate of sampling variability when jackknifing rotation groups was noticeably smaller than when jackknifing random groups, consistent with expectations for the stratified CPS design. This suggests that individual rotation groups are more representative of the population than random subsamples of similar size.

In the third set of experiments, we considered the absolute and relative importance of uncertainty due to calibration of AFDC totals in addition to sampling and stochastic simulation uncertainty. We estimated the percent change in benefits using reform 1, in model runs in which all 10 MATH-CPS model modules were run. Our conclusions from these experiments are:

1. The difference between the standard calibration method for AFDC totals and our method which re-weights households was a much more important contributor to total posterior variance than either sampling or stochastic variability for the reform we used.

2. The particular way in which the households are re-weighted (in our case, by dividing households into two groups or into three groups) had a substantial impact on the estimate of the calibration effect, and therefore also on the estimated contribution of calibration to total posterior variance.

3. If calibration is done based on results from one run of the MATH-CPS model, it is important that the one run be representative of other runs that could be done. Representativeness can be checked by comparing results using several different seeds.

4. A comparison of results from calibrating each run separately to results when calibration is done based on one representative run shows no major reduction in variability when calibration is done for each run.

5. The estimand of interest, the percent change in food stamp benefits under the reform, is only indirectly affected by the number of participants and benefits received in AFDC. Had our estimand been more closely related to the levels that were calibrated, calibration of each run could have given a reduction in total variability.

Table of Contents

6. The estimated calibration effect was similar whether jackknifing rotation groups or jackknifing random groups. This is not surprising since we expect the effect of up-weighting AFDC-eligible households on food stamp benefits to be similar across groups of households.

In our fourth and final set of experiments, we compared the effects of different reforms on the percent change in benefits. In these experiments we incorporated five sources of uncertainty: sampling, stochastic simulation,

income to months. Our conclusions from these experiments are:

1. The absolute and relative importance of each source of uncertainty varied substantially among the reforms we examined.

2. Of the three fixed effects we considered, uncertainty about the unemployment rate and about method of calibrating AFDC totals were more important sources of uncertainty than was the method of allocating annual earned income to months.

3. For the reforms we considered which involved a large change in earnings deduction for many households, uncertainty about the unemployment

microsimulation model, it would be advisable to perform an experiment such as we conducted, on the particular estimand, reform, and microsimulation model of interest.

10 References

Atrostic, B.K., and Bilheimer, Linda (1993), "Modeling the Baseline Distribution of Health Services Spending and Payment", *ASA Proceedings of the Social Statistics Section*, 54-59.

Atrostic, B.K. (1994), "A Multiple Imputation Approach to Microsimulation", *ASA Proceedings of the Section on Survey Research Methods*, 529-534.

Atrostic, B.K. (1995), "Building Sensitivity Tests in a Health Care Microsimulation Model", *ASA Proceedings of the Social Statistics Section*,

Beebout, Harold, and Haworth, Lauren (1989), "Microsimulation Estimates Versus Reality: A Case Study". Unpublished memorandum, Mathematica Policy Research, Inc., Washington, D.C.

Citro, Constance F. and Hanushek, Eric A., (editors) (1991), *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, National Academy Press, Washington, D.C.

Cochran, William C. and Cox, Gertrude M. (1950), *Experimental Designs*, John Wiley and Sons, Inc., New York

Cohen, Michael L. (1991a), "Variance Estimation of Microsimulation Models Through Sample Reuse". In C.F. Citro and E.A. Hanushek (eds.), *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling* (Vol. II). National Academy Press, Washington, D.C.

Cohen, Michael L. (1991b), "Evaluations of Microsimulation Models: Literature Review". In C.F. Citro and E.A. Hanushek (eds.), *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling* (Vol. II). National Academy Press, Washington, D.C.

Cohen, Michael L., Billard, Lynne, Betson, David M., and Ericksen, Eugene P. (1991), "A Validation Experiment with TRIM2". In C.F. Citro and E.A. Hanushek (eds.), *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling* (Vol. II). National Academy Press, Washington, D.C.

Doyle, Pat and Trippe, Carole (1989), " Validation of the Food Stamp Microsimulation Model", Final report to the Food and Nutrition Service, U.S. Department of Agriculture, Mathematica Policy Research Inc., Washington, D.C.

Doyle, Pat and Farley, Dean (1994) "Alternative Strategies for Imputing Premiums and Predicting Expenditures Under Health Care Reform", *ASA Proceedings of the Section on Survey Research Methods*, 517-522.

Farley, Dean E. and Doyle, Pat (1995), "The Effect of Alternative Stochastic Models in Estimating Health Insurance Premiums and Expenditures", *ASA Proceedings of the Social Statistics Section*, 157-162.

Hammersley, J.M. and Handscomb, D.C. (1964), *Monte Carlo Methods*, London: Chapman and Hall.

Heeringa, S., Connor, J., and Woodburn, R.L. (1994), "The 1989 Survey of Consumer Finances, Sample Design Documentation", working paper, ISR, University of Michigan.

Kennickell, Arthur B., and McManus, Douglas A. (1994), "Multiple Imputation of the 1983 and 1989 Waves of the SCF", *ASA Proceedings of the Section on Survey Research Methods*, 523-528.

Martini, Alberto (1991), "Evaluating the Food Stamp Participation Algorithms Used in Microsimulation Models", Technical Working Paper, Mathematica Policy Research, Inc., Washington, DC.

Mathematica Policy Research, Inc. (1994a), "Creation of the 1996 MATH PC Microsimulation Model and Database", Technical Working Paper, Mathematica Policy Research, Inc., Washington, DC.

Mathematica Policy Research, Inc. (1994b), "1996 MATH PC Codebook", Mathematica Policy Research, Inc., Washington, DC.

Mathematica Policy Research, Inc. (1995), "1996 MATH PC Programmer's Guide and Technical Description", Mathematica Policy Research, Inc., Washington, DC.

Maxwell, Scott E. and Delaney, Harold D. (1990), *Designing Experiments and Analyzing Data*, Wadsworth, Belmont CA.

Pudney, Stephen and Sutherland, Holly (1994), "Statistical Reliability and Microsimulation: The Role of Sampling, Simulation and Estimation Errors" , Discussion Paper Series number MU9402, Department of Applied Economics, University of Cambridge, England.

Rubin, Donald (1986), "Statistical Matching Using File Concatenation With Adjusted Weights and Multiple Imputations", *Journal of Business and Economic Statistics*, 87-94.

Table of Contents

Rubin, Donald (1987), *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.

Snedecor, George W. and Cochran, William G. (1980), *Statistical Methods*, Iowa State University Press, Ames IA.

Super, David A., Parrott, Sharon, Steinmetz, Susan and Mann, Cindy, (1996), "The New Welfare Law", Center for Budget and Policy Priorities, Washington, D.C.

U.S. Department of Commerce, Bureau of the Census (1978), "The Current Population Survey: Design and Methodology", Technical Paper number 40.

U.S. Department of Commerce, Bureau of the Census (1992), "Measuring the Effect of Benefits and Taxes on Income and Poverty: 1979 to 1991". Current Population Reports, Series P-60, no. 182-RD, U.S. Government Printing Office, Washington, DC.

U.S. Department of Commerce, Data User Services Division, Data Access and Use Branch, Bureau of the Census (1993a), "Current Population Survey, March 1993 Technical Documentation", The Bureau, Washington, DC.

U.S. Department of Commerce, Bureau of the Census (1993b), "Population Projections of the United States by Age, Sex, and Hispanic Origin: 1993-2050". Current Population Reports, P-25-1104, U.S. Government Printing Office, Washington, DC.

U.S. House of Representatives, Committee on Ways and Means, (1994), "1994 Green Book: Overview of Entitlement Programs", U.S. Government Printing Office, Washington, D.C.

Wolter, Kirk M. (1985), *Introduction to Variance Estimation*, Springer Verlag

Zaslavsky, Alan (1988), "Representing Local Area Adjustments by Reweighting of Households", *Survey Methodology*, Vol 14 No. 2, 265-288.

Zaslavsky, A. and Thurston, S. (1994), "Error Analysis of Food Stamp Microsimulation Models", *ASA Proceedings of the Section on Survey Research Methods*, 535-540.

Zaslavsky, A. and Thurston, S. (1995), "Error Analysis of Food Stamp Microsimulation Models: Further Results", *ASA Proceedings of the Social Statistics Section*, 151-156.

Table of Contents

A Appendices

A.1 Figures

Figure 1: Two-cell calibration weighting factors for AFDC-Basic.

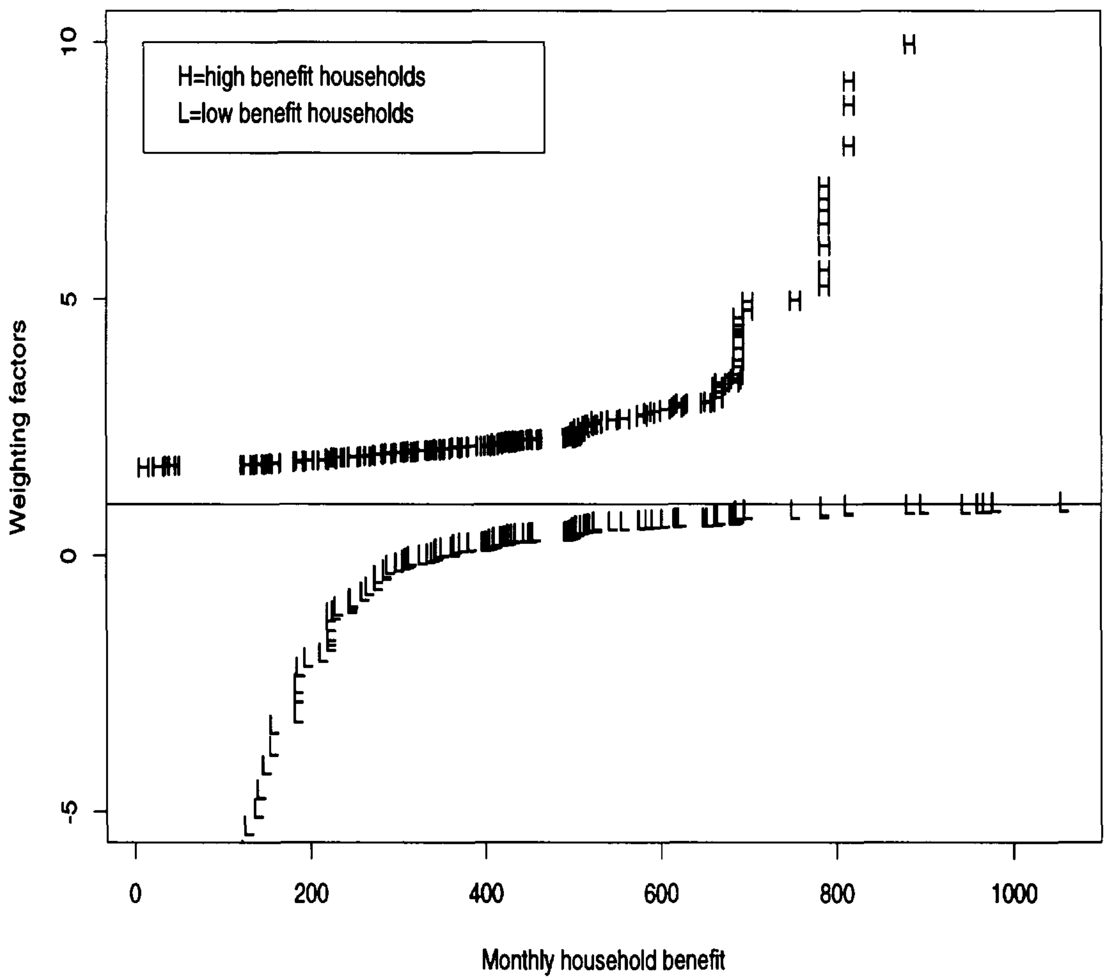

Table of Contents

Figure 2: Two-cell calibration weighting factors for AFDC-UP.

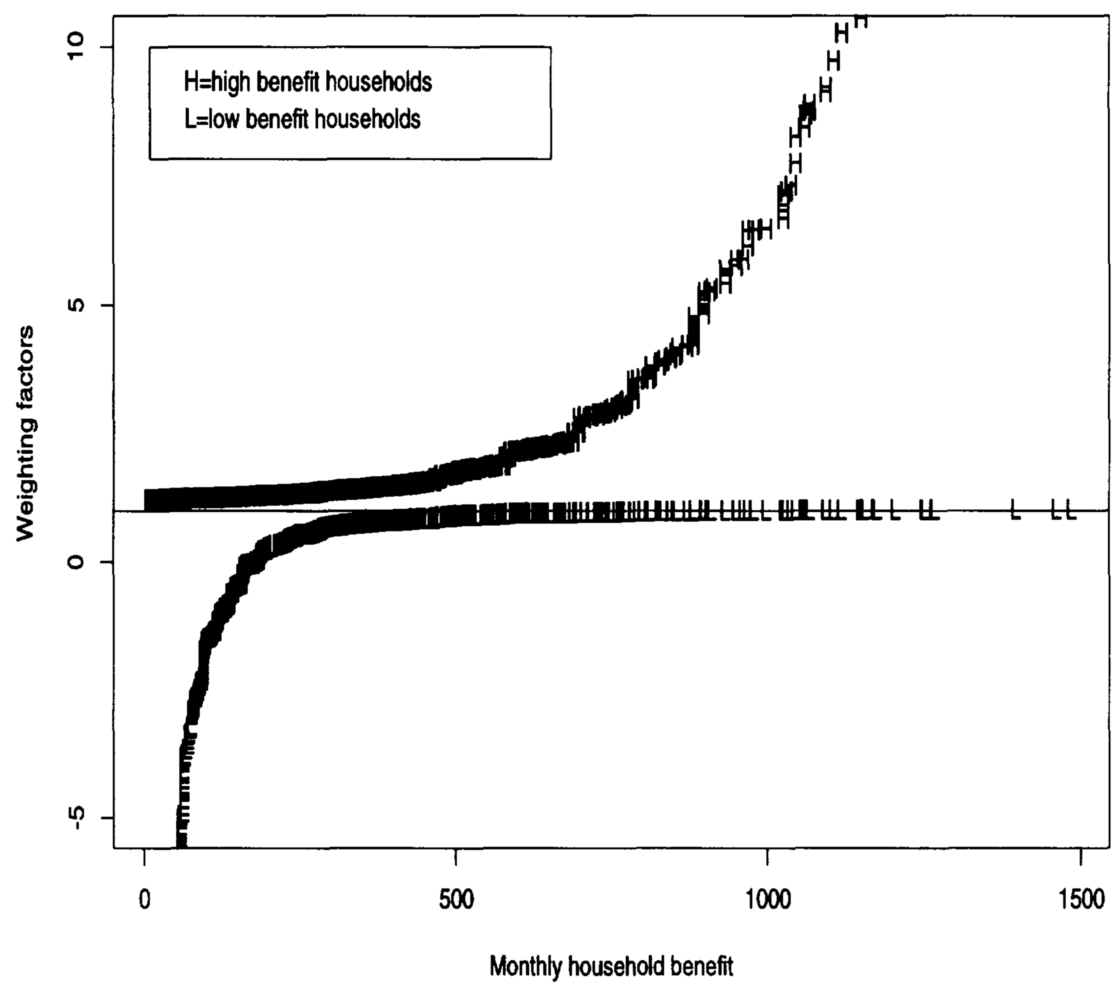

Table of Contents

Figure 3: Comparison of calibration effect, when calibration is done once using MPR seed, to calibration of each run.

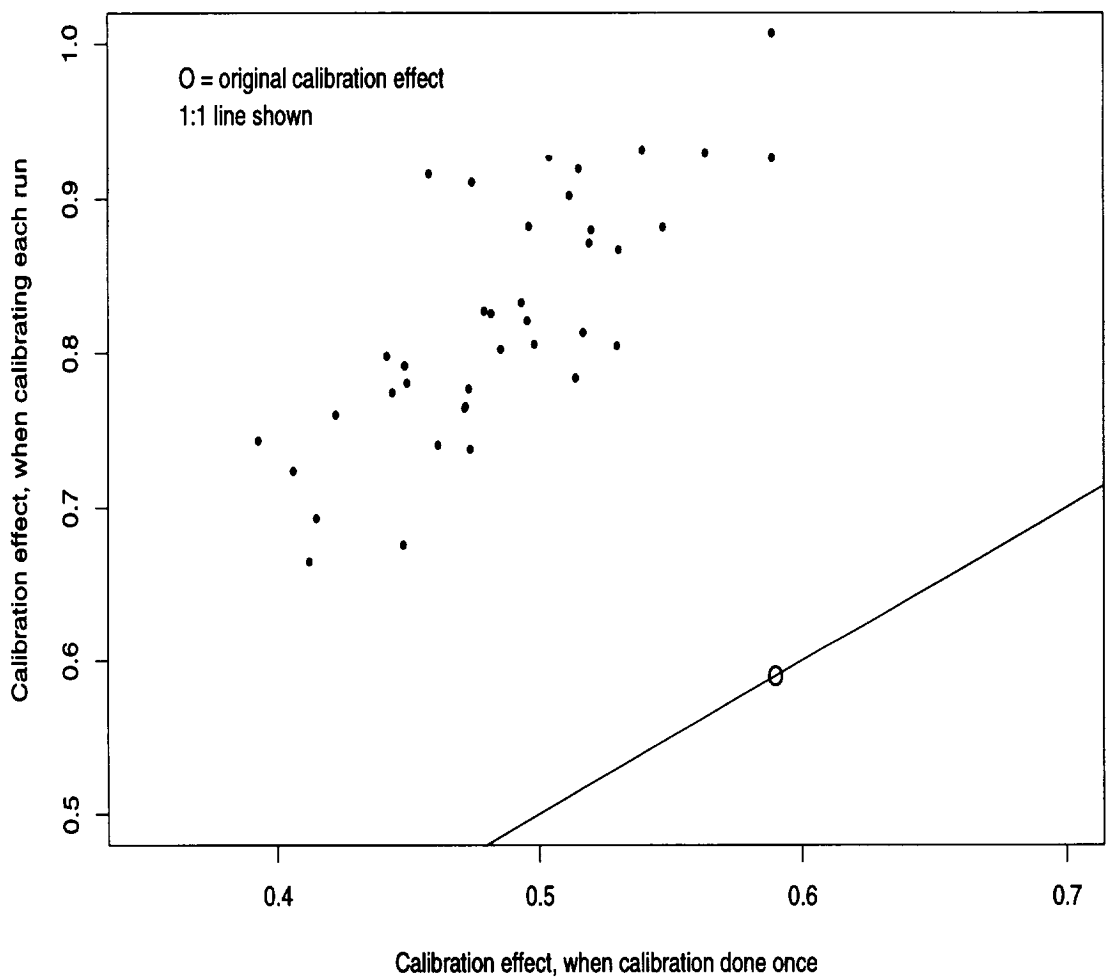

Table of Contents

Figure 4: Comparison of calibration effect, when calibration is done once using weighting factors based on the median of 80 runs, to calibration of each run.

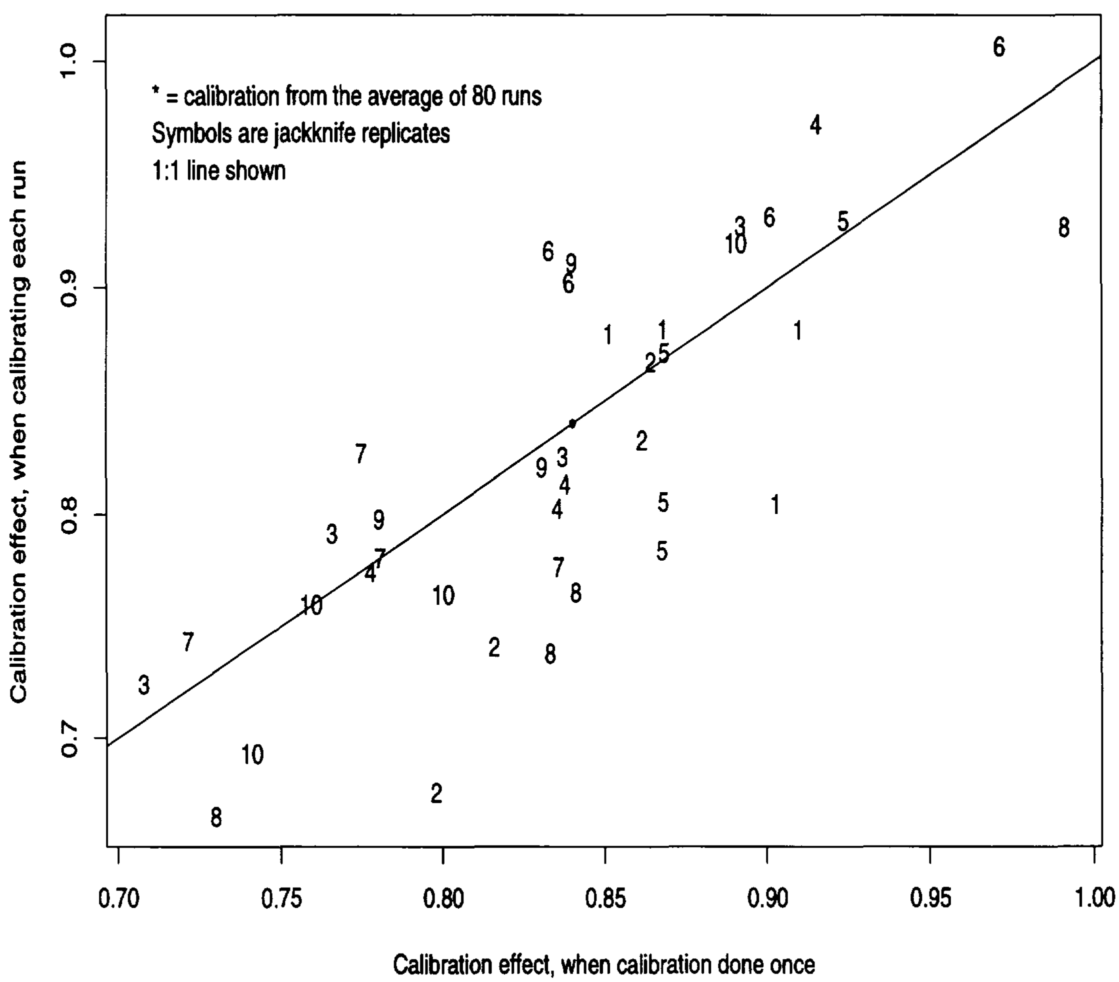

Table of Contents

A.2 ANOVA tables, expected mean squares, and variance estimates

Table 28: Sums of squares from ANOVA using a nested design, for experiments with sampling, and stochastic simulation uncertainty.

ANOVA table

Source	df	SS
Jackknife	$J-1$	$S\sum_j(y_{j.}-y_{..})^2$
Seed in jackknife	$J(S-1)$	$\sum_j\sum_s(y_{js}-y_{j.})^2$

Table 29: Expected mean squares from ANOVA using a nested design, for experiments with sampling and stochastic simulation uncertainty.

Source	E(MS)
Jackknife	$\frac{J}{n(J-1)^2}((J-1)\sigma_s^2+S\sigma_h^2)$
Seed	$\frac{J}{n(J-1)}\sigma_s^2$

Table 30: Variance components estimators from a nested design, for experiments with sampling and stochastic uncertainty.

Variance Component		Estimate
$\hat{\sigma}_h^2$	$=$	$\frac{1}{S}\left\{\frac{n(J-1)^2}{J}\mathrm{MS(jackknife)}-(J-1)\hat{\sigma}_s^2\right\}$
	$=$	$\frac{n(J-1)^2}{JS}\{\mathrm{MS(jackknife)}-\mathrm{MS(seed)}\}$
$\hat{\sigma}_s^2$	$=$	$\frac{n(J-1)}{J}\mathrm{MS(seed)}$

Table of Contents

Table 31: Sums of squares from ANOVA using a crossed design, for experiments with sampling, and stochastic simulation uncertainty.

ANOVA table

Source	df	SS
Jackknife	$J-1$	$S\sum_j(y_{j.}-y_{..})^2$
Seed	$S-1$	$J\sum_s(y_{.s}-y_{..})^2$
Jackknife × seed	$(J-1)(S-1)$	$\sum_j\sum_s(y_{js}-\hat{y}_{js})^2$
		$=\sum_j\sum_s(y_{js}-y_{j.}-y_{.s}+y_{..})^2$

Table 32: Expected mean squares from ANOVA using a crossed design, for experiments with sampling and stochastic simulation uncertainty.

Source	E(MS)
Jackknife	$\frac{J}{n(J-1)^2}(\sigma_s^2+S\sigma_h^2)$
Seed	$\frac{J}{n}\sigma_s^2$
Jackknife × seed	$\frac{J}{n(J-1)^2}\sigma_s^2$

Table 33: Variance components estimators from a crossed design, for experiments with sampling and stochastic uncertainty.

Variance Component		Estimate
$\hat{\sigma}_h^2$	$=$	$\frac{1}{S}\left\{\frac{n(J-1)^2}{J}\mathrm{MS(jackknife)}-\hat{\sigma}_s^2\right\}$
$\hat{\sigma}_{s,1}^2$	$=$	$\frac{n}{J}\mathrm{MS(seed)}$
$\hat{\sigma}_{s,2}^2$	$=$	$\frac{n(J-1)^2}{J}\mathrm{MS(JK \times seed)}$

Table 34: Sums of squares from ANOVA for experiments with sampling, stochastic simulation, and calibration uncertainty. The design is seed nested in jackknife, crossed with calibration.

ANOVA table

Source	df	SS
Jackknife	$J-1$	$2S\sum_j(y_{j..} - y_{...})^2$
Seed in jackknife	$J(S-1)$	$2\sum_j\sum_s(y_{js.} - y_{j..})^2$
Calibration	1	$JS\sum_a(y_{..a} - y_{...})^2$
Calibration × jackknife	$J-1$	$S\sum_j\sum_a(y_{j.a} - \hat{y}_{j.a})^2$
Calibration × seed	$J(S-1)$	$\sum_j\sum_s\sum_a(y_{jsa} - \hat{y}_{jsa})^2$

Table 35: Expected mean squares from ANOVA for experiments with sampling, stochastic simulation, and calibration uncertainty. The design is seed nested in jackknife, crossed with calibration.

ANOVA table

Source	E(MS)
Jackknife	$\frac{2J}{n(J-1)^2}((J-1)\sigma_s^2 + S\sigma_h^2)$
Seed	$\frac{2J}{n(J-1)}\sigma_s^2$
Calibration	$\frac{2J}{n(J-1)}(\sigma_{sc}^2 + S(J-1)\sigma_{hc}^2 + nS(J-1)\sigma_c^2)$
Calibration × jackknife	$\frac{2J}{n(J-1)^2}((J-1)\sigma_{sc}^2 + S\sigma_{hc}^2)$
Calibration × seed	$\frac{2J}{n(J-1)}\sigma_{sc}^2$

Table of Contents

Table 36: Variance components estimators for experiments with sampling, stochastic simulation, and calibration uncertainty. The design is seed nested in jackknife, crossed with calibration.

Variance Component		Estimate
$\hat{\sigma}_h^2$	=	$\frac{n(J-1)^2}{2JS}\{\mathrm{MS(jackknife)} - \mathrm{MS(seed)}\}$
$\hat{\sigma}_s^2$	=	$\frac{n(J-1)}{2J}\mathrm{MS(seed)}$
$\hat{\sigma}_{hc}^2$	=	$\frac{n(J-1)^2}{2JS}\{\mathrm{MS(calib} \times \mathrm{jackknife)} - \mathrm{MS(calib} \times \mathrm{seed)}\}$
$\hat{\sigma}_{sc}^2$	=	$\frac{n(J-1)}{2J}\mathrm{MS(calib} \times \mathrm{seed)}$

Table of Contents

Table 37: Sums of squares from ANOVA for experiments with sampling, stochastic simulation, calibration, unemployment, and income allocation uncertainty. The design is seed nested in jackknife, crossed with calibration crossed with unemployment crossed with income allocation method.

ANOVA table

Source	df	SS
Jackknife (J)	$J-1$	$8S\sum_j(y_{j....}-y_{.....})^2$
Seed in jackknife (S)	$J(S-1)$	$8\sum_j\sum_s(y_{js...}-y_{j....})^2$
Calibration (C)	1	$4JS\sum_c(y_{..c..}-y_{.....})^2$
C × J	$J-1$	$4S\sum_j\sum_c(y_{j.c..}-\hat{y}_{j.c..})^2$
C × S	$J(S-1)$	$4\sum_j\sum_s\sum_c(y_{jsc..}-\hat{y}_{jsc..})^2$
Unemployment (E)	1	$4JS\sum_e(y_{...e.}-y_{.....})^2$
E × J	$J-1$	$4S\sum_j\sum_c(y_{j..e.}-\hat{y}_{j..e.})^2$
E × S	$J(S-1)$	$4\sum_j\sum_s\sum_c(y_{js.e.}-\hat{y}_{js.e.})^2$
Income (I)	1	$4JS\sum_i(y_{....i}-y_{.....})^2$
I × J	$J-1$	$4S\sum_j\sum_c(y_{j...i}-\hat{y}_{j...i})^2$
I × S	$J(S-1)$	$4\sum_j\sum_s\sum_c(y_{js..i}-\hat{y}_{js..i})^2$
C × E	1	$2JS\sum_c\sum_e(y_{..ce.}-\hat{y}_{..ce.})^2$
C × E × J	$J-1$	$2S\sum_j\sum_c\sum_e(y_{j.ce.}-\hat{y}_{j.ce.})^2$
C × E × S	$J(S-1)$	$2\sum_j\sum_s\sum_c\sum_e(y_{jsce.}-\hat{y}_{jsce.})^2$
C × I	1	$2JS\sum_c\sum_i(y_{..c.i}-\hat{y}_{..c.i})^2$
C × I × J	$J-1$	$2S\sum_j\sum_c\sum_i(y_{j.c.i}-\hat{y}_{j.c.i})^2$
C × I × S	$J(S-1)$	$2\sum_j\sum_s\sum_c\sum_i(y_{jsc.i}-\hat{y}_{jsc.i})^2$
E × I	1	$2JS\sum_e\sum_i(y_{...ei}-\hat{y}_{...ei})^2$
E × I × J	$J-1$	$2S\sum_j\sum_e\sum_i(y_{j..ei}-\hat{y}_{j..ei})^2$
E × I × S	$J(S-1)$	$2\sum_j\sum_s\sum_e\sum_i(y_{jsc.i}-\hat{y}_{jsc.i})^2$
C × E × I	1	$JS\sum_c\sum_e\sum_i(y_{..cei}-\hat{y}_{..cei})^2$
C × E × I × J	$J-1$	$S\sum_j\sum_c\sum_e\sum_i(y_{j.cei}-\hat{y}_{j.cei})^2$
C × E × I × S	$J(S-1)$	$\sum_j\sum_s\sum_c\sum_e\sum_i(y_{jscei}-\hat{y}_{jscei})^2$

Table of Contents

Table 38: Expected mean squares from ANOVA for experiments with sampling, stochastic simulation, calibration, unemployment, and income allocation uncertainty. The design is seed nested in jackknife, crossed with calibration crossed with unemployment crossed with income allocation method.

ANOVA table

Source	E(MS)
Jackknife (J)	$\frac{8J}{n(J-1)^2}((J-1)\sigma_s^2 + S\sigma_h^2)$
Seed (S)	$\frac{8J}{n(J-1)}\sigma_s^2$
Calibration (C)	$\frac{8J}{n(J-1)}(\sigma_{sc}^2 + S(J-1)\sigma_{hc}^2 + nS(J-1)\sigma_c^2)$
C × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{sc}^2 + S\sigma_{hc}^2)$
C × S	$\frac{8J}{n(J-1)}\sigma_{sc}^2$
Unemployment (E)	$\frac{8J}{n(J-1)}(\sigma_{su}^2 + S(J-1)\sigma_{he}^2 + nS(J-1)\sigma_e^2)$
E × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{se}^2 + S\sigma_{he}^2)$
E × S	$\frac{8J}{n(J-1)}\sigma_{se}^2$
Income (I)	$\frac{8J}{n(J-1)}(\sigma_{si}^2 + S(J-1)\sigma_{hi}^2 + nS(J-1)\sigma_i^2)$
I × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{si}^2 + S\sigma_{hi}^2)$
I × S	$\frac{8J}{n(J-1)}\sigma_{si}^2$
C × E	$\frac{8J}{n(J-1)}(\sigma_{sce}^2 + S(J-1)\sigma_{hce}^2 + nS(J-1)\sigma_{ce}^2)$
C × E × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{sce}^2 + S\sigma_{hce}^2)$
C × E × S	$\frac{8J}{n(J-1)}\sigma_{sce}^2$
C × I	$\frac{8J}{n(J-1)}(\sigma_{sci}^2 + S(J-1)\sigma_{hci}^2 + nS(J-1)\sigma_{ci}^2)$
C × I × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{sci}^2 + S\sigma_{hci}^2)$
C × I × S	$\frac{8J}{n(J-1)}\sigma_{sci}^2$
E × I	$\frac{8J}{n(J-1)}(\sigma_{sei}^2 + S(J-1)\sigma_{hei}^2 + nS(J-1)\sigma_{ei}^2)$
E × I × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{sei}^2 + S\sigma_{hei}^2)$
E × I × S	$\frac{8J}{n(J-1)}\sigma_{sei}^2$
C × E × I	$\frac{8J}{n(J-1)}(\sigma_{scei}^2 + S(J-1)\sigma_{hcei}^2 + nS(J-1)\sigma_{cei}^2)$
C × E × I × J	$\frac{8J}{n(J-1)^2}((J-1)\sigma_{scei}^2 + S\sigma_{hcei}^2)$
C × E × I × S	$\frac{8J}{n(J-1)}\sigma_{scei}^2$

Table of Contents

Table 39: Variance components estimators for experiments with sampling, stochastic simulation, calibration, unemployment and income allocation uncertainty. The design is seed nested in jackknife, crossed with calibration crossed with unemployment crossed with income allocation method.

Variance Component		Estimate
$\hat{\sigma}^2_h$	=	$\frac{n(J-1)^2}{8JS}$ {MS(J) − MS(S)}
$\hat{\sigma}^2_s$	=	$\frac{n(J-1)}{8J}$MS(S)
$\hat{\sigma}^2_{hc}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(C × J) − MS(C × S)}
$\hat{\sigma}^2_{sc}$	=	$\frac{n(J-1)}{8J}$MS(C × S)
$\hat{\sigma}^2_{he}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(E × J) − MS(E × S)}
$\hat{\sigma}^2_{se}$	=	$\frac{n(J-1)}{8J}$MS(E × S)
$\hat{\sigma}^2_{hi}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(I × J) − MS(I × S)}
$\hat{\sigma}^2_{si}$	=	$\frac{n(J-1)}{8J}$MS(I × S)
$\hat{\sigma}^2_{hce}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(C × E × J) − MS(C × E × S)}
$\hat{\sigma}^2_{sce}$	=	$\frac{n(J-1)}{8J}$MS(C × E × S)
$\hat{\sigma}^2_{hci}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(C × I × J) − MS(C × I × S)}
$\hat{\sigma}^2_{sci}$	=	$\frac{n(J-1)}{8J}$MS(C × I × S)
$\hat{\sigma}^2_{hei}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(E × I × J) − MS(E × I × S)}
$\hat{\sigma}^2_{sei}$	=	$\frac{n(J-1)}{8J}$MS(E × I × S)
$\hat{\sigma}^2_{hcei}$	=	$\frac{n(J-1)^2}{8JS}$ {MS(C × E × I × J) − MS(C × E × I × S)}
$\hat{\sigma}^2_{scei}$	=	$\frac{n(J-1)}{8J}$MS(C × E × I × S)

Table of Contents

A.3 Tables of results from individual experiments

A.3.1 Experiments using fewer than 10 MATH-CPS model modules

Table 40: Results from running just the FSTAMP reform part of the MATH-CPS model incorporating sampling and stochastic variability, using a jackknife of 20 random groups and 4 seeds nested under reform 1.

	full run	single run
$\hat{\beta}$	9.827	
$\sqrt{\sigma_h^2/n}$	0.805	
$\sqrt{\sigma_s^2/n}$	0.233	
SE($\beta \mid \hat{\beta}$)	.805	.838
95% interval	(8.217, 11.438)	(8.152, 11.503)
% var due to:		
sampling	99.889	92.237
seed	0.111	7.763

Table 41: Results from running the FSTAMP baselaw and reform parts of the MATH-CPS model incorporating sampling and stochastic variability, using a jackknife of 20 random groups and 4 seeds nested under reform 1.

	full run	single run
$\hat{\beta}$	9.738	
$\sqrt{\sigma_h^2/n}$	0.699	
$\sqrt{\sigma_s^2/n}$	0.218	
SE($\beta \mid \hat{\beta}$)	.700	0.732
95% interval	(8.339, 11.137)	(8.274, 11.203)
% var due to:		
sampling	99.872	91.106
seed	0.128	8.894

Table of Contents

Table 42: Results from running 8 modules of the MATH-CPS model incorporating sampling and stochastic variability, using a jackknife of 20 random groups and 4 seeds nested under reform 1.

	full run	single run
$\hat{\beta}$	9.932	
$\sqrt{\sigma_h^2/n}$	0.479	
$\sqrt{\sigma_s^2/n}$	0.223	
SE($\beta \mid \hat{\beta}$)	.479	.528
95% interval	(8.973, 10.891)	(8.876, 10.988)
% var due to:		
sampling	99.716	82.206
seed	0.284	17.794

Table 43: Results from running 9 modules of the MATH-CPS model incorporating sampling and stochastic variability, using a jackknife of 20 random groups and 4 seeds nested under reform 1.

	full run	single run
$\hat{\beta}$	9.234	
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.714	
SE($\beta \mid \hat{\beta}$)	.082	0.714
95% interval	(9.071, 9.398)	(7.807, 10.662)
% var due to:		
sampling	0.0	0.0
seed	100.0	100.0

Table of Contents

A.3.2 Experiments with sampling and stochastic uncertainty

Table 44: Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds nested, seed set A, and reform 1.

	full run	single run
$\hat{\beta}$	9.215	
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.709	
SE($\beta \mid \hat{\beta}$)	.081	.709
95% interval	(9.053, 9.378)	(7.800, 10.634)
% var due to:		
sampling	0.0	0.0
seed	100.0	100.0

Table 45: Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds nested, seed set B, under reform 1

	full run	single run
$\hat{\beta}$	9.118	
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.771	
SE($\beta \mid \hat{\beta}$)	.088	.771
95% interval	(8.941, 9.295)	(7.576, 10.661)
% var due to:		
sampling	0.0	0.0
seed	100.0	100.0

Table of Contents

Table 46: Results from running 10 modules of the MATH-CPS model, using a jackknife of 19 random groups, 4 seeds nested, seed set A, and reform 1.

	full run	single run
$\hat{\beta}$	9.256	
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.716	
$SE(\beta \mid \hat{\beta})$	.084	.716
95% interval	(9.087, 9.425)	(7.824, 10.688)
% var due to:		
sampling	0.0	0.0
seed	100.0	100.0

Table 47: Results from running 10 modules of the MATH-CPS model, using a jackknife of 19 random groups, 4 seeds nested, seed set B, under reform 1.

	full run	single run
$\hat{\beta}$	9.080	
$\sqrt{\sigma_h^2/n}$	0.561	
$\sqrt{\sigma_s^2/n}$	0.750	
$SE(\beta \mid \hat{\beta})$	.568	.936
95% interval	(7.945, 10.215)	(7.208, 10.953)
% var due to:		
sampling	97.575	35.85
seed	2.425	64.15

Table 48: Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds crossed, seed set A, under reform 1.

	full run		single run	
	using MS(seed)	using MS(JK*seed)	using MS(seed)	using MS(JK*seed)
$\hat{\beta}$	9.609	9.609		
$\sqrt{\sigma_h^2/n}$	0.525	0.455		
$\sqrt{\sigma_s^2/n}$	0.423	0.674		
$SE(\beta \mid \hat{\beta})$	0.527	0.461	.674	.813
95% interval	(8.554, 10.664)	(8.686, 10.532)	(8.260, 10.957)	(7.983, 11.235)
% var due to:				
sampling	99.154	97.195	60.665	31.317
seed	0.846	2.805	39.335	68.683

Table of Contents

Table 49: Results from running 10 modules of the MATH-CPS model, using a jackknife of 20 random groups, 4 seeds crossed, seed set B, and reform 1.

	full run		single run	
	using MS(seed)	using MS(JK*seed)	using MS(seed)	using MS(JK*seed)
$\hat{\beta}$	8.876	8.876		
$\sqrt{\sigma_h^2/n}$	0.0	0.288		
$\sqrt{\sigma_s^2/n}$	1.026	0.713		
SE($\beta \mid \hat{\beta}$)	0.118	0.299	1.026	.769
95% interval	(8.641, 9.112)	(8.278, 9.475)	(6.824, 10.929)	(7.338, 10.415)
% var due to:				
sampling	0	92.535	0	14.023
seed	100	7.465	100	85.977

Table 50: Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds nested, seed set A, and reform 1.

	full run	single run
$\hat{\beta}$	9.273	
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.715	
SE($\beta \mid \hat{\beta}$)	.085	.715
95% interval	(9.102, 9.444)	(7.842, 10.703)
% var due to:		
sampling	0.0	0.0
seed	100.0	100.0

Table 51: Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds nested, seed set B, and reform 1.

	full run	single run
$\hat{\beta}$	9.131	
$\sqrt{\sigma_h^2/n}$	0.249	
$\sqrt{\sigma_s^2/n}$	0.734	
SE($\beta \mid \hat{\beta}$)	.264	.775
95% interval	(8.603, 9.658)	(7.580, 10.681)
% var due to:		
sampling	88.935	10.3
seed	11.065	89.7

Table of Contents

Table 52: Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds crossed, seed set A, and reform 1.

	full run		single run	
	using MS(seed)	using MS(JK*seed)	using MS(seed)	using MS(JK*seed)
$\hat{\beta}$	9.507	9.507		
$\sqrt{\sigma_h^2/n}$	0.167	0.164		
$\sqrt{\sigma_s^2/n}$	0.646	0.655		
SE($\beta \mid \hat{\beta}$)	0.184	0.181	.667	.675
95% interval	(9.139, 9.875)	(9.144, 9.870)	(8.173, 10.841)	(8.157, 10.857)
% var due to:				
sampling	82.415	81.366	6.275	5.871
seed	17.585	18.634	93.725	94.129

Table 53: Results from running 10 modules of the MATH-CPS model, using a jackknife of 8 rotation groups, 10 seeds crossed, seed set B, and reform 1

	full run		single run	
	using MS(seed)	using MS(JK*seed)	using MS(seed)	using MS(JK*seed)
$\hat{\beta}$	8.942	8.942		
$\sqrt{\sigma_h^2/n}$	0.118	0.165		
$\sqrt{\sigma_s^2/n}$	0.841	0.758		
SE($\beta \mid \hat{\beta}$)	0.155	0.188	0.849	0.776
95% interval	(8.631, 9.253)	(8.565, 9.319)	(7.244, 10.640)	(7.391, 10.493)
% var due to:				
sampling	58.15	76.902	1.946	4.54
seed	41.85	23.098	98.054	95.46

Table of Contents

A.3.3 Experiments with sampling, stochastic and calibration uncertainty

Table 54: MATH-CPS model results, calibrating AFDC totals based on weighting factors from the MPR seed, separating households into 2 groups. A jackknife of 10 random groups was used with 4 seeds nested. The experiment uses seed set A, and is run under reform 1.

	full run	single run
$\hat{\beta}$	8.954 (.349)	8.954 (.684)
$\hat{\beta}_C$	-0.329 (.020)	-0.329 (.119)
$\sqrt{\sigma_h^2/n}$	0.334	
$\sqrt{\sigma_s^2/n}$	0.596	
SE$(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$	.480	.768
95% interval	(7.994, 9.915)	(7.418, 10.490)
% var due to:		
sampling	48.482	18.954
seed	4.281	60.253
calib	47.067	18.400
calib × sample	0.0	0.0
calib × seed	0.170	2.393

Table of Contents

Table 55: MATH-CPS model results, calibrating AFDC totals based weighting factors from the MPR seed, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested. The experiment uses seed set A, and is run under reform 1.

	full run	single run
$\hat{\beta}$	8.816 (.105)	8.816 (.632)
$\hat{\beta}_C$	-0.488 (0.052)	-0.488 (.066)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.632	
SE($\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C$)	0.502	0.801
95% interval	(7.812, 9.821)	(7.214, 10.419)
% var due to:		
sampling	0.00	0.00
seed	4.401	62.208
calib	94.507	37.106
calib × sample	1.074	0.421
calib × seed	0.019	0.265

Table 56: MATH-CPS model results, calibrating AFDC totals on each run, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested. The experiment uses seed set A, and is run under reform 1.

	full run	single run
$\hat{\beta}$	8.475 (.101)	8.475 (.606)
$\hat{\beta}_C$	-0.829 (.105)	-0.829 (.127)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.606	
SE($\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C$)	0.842	1.035
95% interval	(6.791, 10.159)	(6.405, 10.545)
% var due to:		
sampling	0.00	0.00
seed	1.437	34.257
calib	97.014	64.242
calib × sample	1.528	1.012
calib × seed	0.021	0.489

Table of Contents

Table 57: MATH-CPS model results, calibrating AFDC totals based on median weighting factors, separating households into 3 groups. A jackknife of 10 random groups was used with 4 seeds nested. The experiment uses seed set A, and is run under reform 1.

	full run	single run
$\hat{\beta}$	8.466 (.103)	8.466 (.615)
$\hat{\beta}_C$	-0.839 (.066)	-0.839 (.087)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.615	
SE$(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$	0.848	1.044
95% interval	(6.771, 10.161)	(6.378, 10.553)
% var due to:		
sampling	0.00	0.00
seed	1.463	34.728
calib	97.933	64.579
calib × sample	0.592	0.390
calib × seed	0.013	0.303

Table 58: MATH-CPS model results, calibrating AFDC totals on each run, separating households into 3 groups. A jackknife of 8 rotation groups was used with 4 seeds nested. The experiment uses seed set A, and is run under reform 1.

	full run	single run
$\hat{\beta}$	8.420 (.292)	8.420 (.646)
$\hat{\beta}_C$	-.900 (.113)	-.900 (.133)
$\sqrt{\sigma_h^2/n}$	0.270	
$\sqrt{\sigma_s^2/n}$	0.587	
SE$(\beta + c\beta_C \mid \hat{\beta}, \hat{\beta}_C)$	0.953	1.116
95% interval	(6.514, 10.327)	(6.188, 10.652)
% var due to:		
sampling	8.018	5.850
seed	1.353	27.634
calib	89.220	65.090
calib × sample	1.389	1.013
calib × seed	0.020	0.413

Table of Contents

A.3.4 Experiments with five sources of uncertainty

certainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	8.468 (.689)	8.468 (.858)
$\hat{\beta}_C$	-.821 (.101)	-.821 (.106)
$\hat{\beta}_E$	-.081 (.357)	-.081 (.366)
$\hat{\beta}_I$	.037 (.126)	.037 (.140)
$\hat{\beta}_{CE}$	.017 (.006)	.017 (.024)
$\hat{\beta}_{CI}$	-.003 (.005)	-.003 (.010)
$\hat{\beta}_{EI}$	.004 (.032)	.004 (.042)
$\hat{\beta}_{CEI}$	.000 (.007)	.000 (.008)
$\sqrt{\sigma_h^2/n}$	.674	
$\sqrt{\sigma_s^2/n}$	.530	
SE(true mean | estimates)	1.145	1.259
95% interval	(6.177, 10.758)	(5.949, 10.986)
% var due to:		
sampling (J)	34.638	28.660
seed (S)	1.529	17.717
calib (C)	51.431	42.555
C × J	0.763	0.632

Table 60: MATH-CPS model results under reform 1 using 5 sources of uncertainty, and seed set B. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	8.203 (.210)	8.203 (.787)
$\hat{\beta}_C$	-.796 (.028)	-.796 (.104)
$\hat{\beta}_E$	-.095 (.028)	-.095 (.104)
$\hat{\beta}_I$	.035 (.019)	.035 (.070)
$\hat{\beta}_{CE}$	.023 (.011)	.023 (.016)
$\hat{\beta}_{CI}$	-.001 (.003)	-.001 (.010)
$\hat{\beta}_{EI}$	.008 (.005)	.008 (.020)
$\hat{\beta}_{CEI}$	-.001 (.002)	-.001 (.002)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.787	
SE(true mean | estimates)	0.832	1.136
95% interval	(6.540, 9.867)	(5.931, 10.476)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	6.392	47.920
calib (C)	91.732	49.120
C × J	0.000	0.000
C × S	0.112	0.842
unemployment (E)	1.319	0.706
E × J	0.000	0.000
E × S	0.112	0.842
Income (I)	0.173	0.093
I × J	0.000	0.000
I × S	0.050	0.374
C × E	0.077	0.041
C × E × J	0.016	0.009
C × E × S	0.001	0.010
C × I	0.000	0.000
C × I × J	0.000	0.000
C × I × S	0.001	0.008
E × I	0.009	0.005
E × I × J	0.000	0.000
E × I × S	0.004	0.031
C × E × I	0.000	0.000
C × E × I × J	0.000	0.000
C × E × I × S	0.000	0.000

Table of Contents

Table 61: MATH-CPS model results under reform 2 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	10.435 (.093)	10.435 (.349)
$\hat{\beta}_C$	-.502 (.053)	-.502 (.061)
$\hat{\beta}_E$	.803 (.024)	.803 (.090)
$\hat{\beta}_I$	-.425 (.107)	-.425 (.129)
$\hat{\beta}_{CE}$	-.026 (.004)	-.026 (.014)
$\hat{\beta}_{CI}$	.027 (.007)	.027 (.010)
$\hat{\beta}_{EI}$	-.033 (.044)	-.033 (.048)
$\hat{\beta}_{CEI}$	.003 (.000)	.003 (.000)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.349	
SE(true mean | estimates)	1.051	1.110
95% interval	(8.333, 12.537)	(8.215, 12.655)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	0.789	9.898
calib (C)	22.766	20.409
C × J	0.249	0.223
C × S	.006	0.077
unemployment (E)	58.354	52.313
E × J	0.000	0.000
E × S	0.052	0.657
Income (I)	16.337	14.646
I × J	1.004	0.900
I × S	0.036	0.458
C × E	0.062	0.055
C × E × J	0.000	0.000
C × E × S	0.001	0.015
C × I	0.065	0.058
C × I × J	0.004	0.004
C × I × S	0.000	0.005
E × I	0.099	0.088
E × I × J	0.171	0.153
E × I × S	0.003	0.036
C × E × I	0.001	0.001
C × E × I × J	0.000	0.000
C × E × I × S	0.000	0.002

Table of Contents

Table 62: MATH-CPS model results under reform 3 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	2.838 (.055)	2.838 (.131)
$\hat{\beta}_C$	.086 (.008)	.086 (.031)
$\hat{\beta}_E$	-.049 (.078)	-.049 (.079)
$\hat{\beta}_I$	-.040 (.008)	-.040 (.031)
$\hat{\beta}_{CE}$	-.008 (.012)	-.008 (.014)
$\hat{\beta}_{CI}$	.003 (.009)	.003 (.009)
$\hat{\beta}_{EI}$	-.002 (.002)	-.002 (.007)
$\hat{\beta}_{CEI}$	.000 (.000)	.000 (.001)
$\sqrt{\sigma_h^2/n}$	0.044	
$\sqrt{\sigma_s^2/n}$	0.124	
SE(true mean | estimates)	0.145	0.193
95% interval	(2.548, 3.128)	(2.452, 3.224)
% var due to:		
sampling (J)	9.133	5.152
seed (S)	5.192	41.005
calib (C)	35.247	19.883
C × J	0.000	0.000
C × S	0.324	2.558
unemployment (E)	11.614	6.551
E × J	28.994	16.356
E × S	0.065	0.515
Income (I)	7.683	4.334
I × J	0.000	0.000
I × S	0.324	2.557
C × E	0.272	0.153
C × E × J	0.710	0.401
C × E × S	0.020	0.156
C × I	0.037	0.021
C × I × J	0.344	0.194
C × I × S	0.003	0.024
E × I	0.023	0.013
E × I × J	0.000	0.000
E × I × S	0.015	0.122
C × E × I	0.000	0.000
C × E × I × J	0.000	0.000
C × E × I × S	0.001	0.005

Table 63: MATH-CPS model results under reform 4 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	5.314 (.157)	5.314 (.587)
$\hat{\beta}_C$	-.355 (.081)	-.355 (.085)
$\hat{\beta}_E$	-.047 (.227)	-.047 (.233)
$\hat{\beta}_I$	-.019 (.124)	-.019 (.137)
$\hat{\beta}_{CE}$	.007 (.003)	.007 (.012)
$\hat{\beta}_{CI}$	.001 (.009)	.001 (.010)
$\hat{\beta}_{EI}$	.002 (.021)	.002 (.024)
$\hat{\beta}_{CEI}$	.000 (.003)	.000 (.004)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.587	
SE(true mean | estimates)	0.476	0.745
95% interval	(4.361, 6.266)	(3.824, 6.803)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	10.857	62.195
calib (C)	55.545	22.729
C × J	2.854	1.168
C × S	0.022	0.126
unemployment (E)	0.956	0.391
E × J	22.516	9.214
E × S	0.100	0.572
Income (I)	0.153	0.063
I × J	6.626	2.711
I × S	0.117	0.670
C × E	0.021	0.009
C × E × J	0.000	0.000
C × E × S	0.005	0.027
C × I	0.000	0.000
C × I × J	0.034	0.014
C × I × S	0.001	0.003
E × I	0.001	0.000
E × I × J	0.182	0.075
E × I × S	0.005	0.031
C × E × I	0.000	0.000
C × E × I × J	0.005	0.002
C × E × I × S	0.000	0.000

Table of Contents

Table 64: MATH-CPS model results under reform 5 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	26.056 (1.783)	26.056 (1.893)
$\hat{\beta}_C$	-1.705 (0.218)	-1.705 (.223)
$\hat{\beta}_E$	-.123 (.532)	-.123 (.556)
$\hat{\beta}_I$	.018 (.245)	.018 (.256)
$\hat{\beta}_{CE}$	.016 (.009)	.016 (.032)
$\hat{\beta}_{CI}$	.001 (.011)	.001 (.017)
$\hat{\beta}_{EI}$	.002 (.011)	.002 (.042)
$\hat{\beta}_{CEI}$	.001 (.007)	.001 (.008)
$\sqrt{\sigma_h^2/n}$	1.774	
$\sqrt{\sigma_s^2/n}$	0.659	
SE(true mean | estimates)	2.548	2.633
95% interval	(20.960, 31.153)	(20.790, 31.322)
% var due to:		
sampling (J)	48.481	45.403
seed (S)	0.477	6.259
calib (C)	44.777	41.935
C × J	0.727	0.681
C × S	0.003	0.036
unemployment (E)	0.233	0.218
E × J	4.335	4.060
E × S	0.031	0.407
Income (I)	0.005	0.005
I × J	0.915	0.857
I × S	0.007	0.090
C × E	0.004	0.004
C × E × J	0.000	0.000
C × E × S	0.001	0.015
C × I	0.000	0.000
C × I × J	0.002	0.001
C × I × S	0.000	0.002
E × I	0.000	0.000
E × I × J	0.000	0.000
E × I × S	0.002	0.026
C × E × I	0.000	0.000
C × E × I × J	0.001	0.001
C × E × I × S	0.000	0.000

Table of Contents

Table 65: MATH-CPS model results under reform 6 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seed nested.

	full run	single run
$\hat{\beta}$	3.081 (.062)	3.081 (.232)
$\hat{\beta}_C$	.070 (.006)	0.070 (.023)
$\hat{\beta}_E$	.279 (.018)	0.279 (.069)
$\hat{\beta}_I$	-.039 (.016)	-.039 (.059)
$\hat{\beta}_{CE}$	.007 (.003)	.007 (.011)
$\hat{\beta}_{CI}$	-.004 (.015)	-.004 (.016)
$\hat{\beta}_{EI}$	.001 (.025)	.001 (.029)
$\hat{\beta}_{CEI}$	.000 (.000)	.000 (.003)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.232	
SE(true mean | estimates)	0.300	0.385
95% interval	(2.482, 3.680)	(2.311, 3.851)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	4.265	36.204
calib (C)	5.492	3.330
C × J	0.000	0.000
C × S	0.041	0.347
unemployment (E)	86.843	52.653
E × J	0.000	0.000
E × S	0.378	3.211
Income (I)	1.696	1.028
I × J	0.000	0.000
I × S	0.279	2.365
C × E	0.055	0.033
C × E × J	0.000	0.000
C × E × S	0.009	0.075
C × I	0.016	0.010
C × I × J	0.243	0.147
C × I × S	0.003	0.021
E × I	0.002	0.001
E × I × J	0.657	0.399
E × I × S	0.019	0.160
C × E × I	0.000	0.000
C × E × I × J	0.000	0.000
C × E × I × S	0.002	0.014

Table of Contents

Table 66: MATH-CPS model results under reform 7 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seeds nested.

	full run	single run
$\hat{\beta}$	.308 (.080)	.308 (.299)
$\hat{\beta}_C$	-.028 (.039)	-.028 (.048)
$\hat{\beta}_E$	-.022 (.021)	-.022 (.080)
$\hat{\beta}_I$	-.002 (.057)	-.002 (.067)
$\hat{\beta}_{CE}$	.002 (.005)	.002 (.017)
$\hat{\beta}_{CI}$	.000 (.003)	.000 (.004)
$\hat{\beta}_{EI}$	.004 (.004)	.004 (.014)
$\hat{\beta}_{CEI}$	.000 (.000)	.000 (.001)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.299	
SE(true mean | estimates)	0.114	0.323
95% interval	(0.081, 0.535)	(-0.338, 0.954)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	49.626	85.709
calib (C)	6.076	0.750
C × J	11.102	1.370
C × S	0.477	0.823
unemployment (E)	3.848	0.475
E × J	0.000	0.000
E × S	3.533	6.101
Income (I)	0.031	0.004
I × J	24.080	2.971
I × S	0.752	1.298
C × E	0.033	0.004
C × E × J	0.000	0.000
C × E × S	0.166	0.286
C × I	0.000	0.000
C × I × J	0.058	0.007
C × I × S	0.006	0.010
E × I	0.109	0.013
E × I × J	0.000	0.000
E × I × S	0.103	0.178
C × E × I	0.001	0.000
C × E × I × J	0.000	0.000
C × E × I × S	0.001	0.002

Table of Contents

Table 67: MATH-CPS model results under reform 8 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seeds nested.

	full run	single run
$\hat{\beta}$	.215 (.130)	.215 (.137)
$\hat{\beta}_C$	.077 (.002)	.077 (.009)
$\hat{\beta}_E$	-.058 (.023)	-.058 (.026)
$\hat{\beta}_I$	.010 (.005)	.010 (.017)
$\hat{\beta}_{CE}$	.002 (.002)	.002 (.003)
$\hat{\beta}_{CI}$	-.001 (.004)	-.001 (.004)
$\hat{\beta}_{EI}$	.000 (.009)	.000 (.009)
$\hat{\beta}_{CEI}$	.000 (.001)	.000 (.001)
$\sqrt{\sigma_h^2/n}$	0.130	
$\sqrt{\sigma_s^2/n}$	0.045	
SE(true mean | estimates)	0.165	0.172
95% interval	(-0.115, 0.544)	(-0.129, 0.558)
% var due to:		
sampling (J)	62.003	57.016
seed (S)	0.537	6.910
calib (C)	22.059	20.284
C × J	0.000	0.000
C × S	0.023	0.291
unemployment (E)	12.493	11.488
E × J	2.001	1.840
E × S	0.033	0.421
Income (I)	0.382	0.351
I × J	0.000	0.000
I × S	0.078	0.999
C × E	0.015	0.014
C × E × J	0.010	0.009
C × E × S	0.002	0.020
C × I	0.002	0.002
C × I × J	0.055	0.050
C × I × S	0.000	0.003
E × I	0.000	0.000
E × I × J	0.302	0.278
E × I × S	0.001	0.015
C × E × I	0.000	0.000
C × E × I × J	0.006	0.005
C × E × I × S	0.000	0.002

Table of Contents

Table 68: MATH-CPS model results under reform 9 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seeds nested.

	full run	single run
$\hat{\beta}$	.302 (.163)	.302 (.168)
$\hat{\beta}_C$	.056 (.012)	.056 (.045)
$\hat{\beta}_E$	.000 (.010)	.000 (.038)
$\hat{\beta}_I$	-.003 (.005)	-.003 (.019)
$\hat{\beta}_{CE}$	.002 (.004)	.002 (.015)
$\hat{\beta}_{CI}$	-.003 (.013)	-.003 (.014)
$\hat{\beta}_{EI}$	-.001 (.009)	-.001 (.010)
$\hat{\beta}_{CEI}$	-.001 (.004)	-.001 (.005)
$\sqrt{\sigma_h^2/n}$	0.163	
$\sqrt{\sigma_s^2/n}$	0.040	
SE(true mean | estimates)	0.174	0.189
95% interval	(-.047, 0.651)	(-0.076, 0.680)
% var due to:		
sampling (J)	87.303	74.448
seed (S)	0.379	4.519
calib (C)	10.480	8.937
C × J	0.000	0.000
C × S	0.465	5.555
unemployment (E)	0.000	0.000
E × J	0.000	0.000
E × S	0.337	4.023
Income (I)	0.034	0.029
I × J	0.000	0.000
I × S	0.082	0.974
C × E	0.012	0.010
C × E × J	0.000	0.000
C × E × S	0.052	0.616
C × I	0.029	0.024
C × I × J	0.517	0.441
C × I × S	0.006	0.073
E × I	0.001	0.001
E × I × J	0.245	0.209
E × I × S	0.007	0.081
C × E × I	0.002	0.002
C × E × I × J	0.050	0.042
C × E × I × S	0.001	0.016

Table 69: MATH-CPS model results under reform 10 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seeds nested.

	full run	single run
$\hat{\beta}$	.166 (.187)	.166 (.241)
$\hat{\beta}_C$	-.057 (.026)	-.057 (.030)
$\hat{\beta}_E$	-.154 (.046)	-.154 (.058)
$\hat{\beta}_I$	-.030 (.071)	-.030 (.090)
$\hat{\beta}_{CE}$	.004 (.006)	.004 (.007)
$\hat{\beta}_{CI}$	.006 (.002)	.006 (.007)
$\hat{\beta}_{EI}$	.009 (.014)	.009 (.020)
$\hat{\beta}_{CEI}$	.001 (.003)	.001 (.004)
$\sqrt{\sigma_h^2/n}$	0.182	
$\sqrt{\sigma_s^2/n}$	0.159	
SE(true mean | estimates)	0.266	0.315
95% interval	(-0.367, 0.699)	(-0.464, 0.796)
% var due to:		
sampling (J)	46.520	33.273
seed (S)	2.538	25.415
calib (C)	4.645	3.322
C × J	0.918	0.656
C × S	0.023	0.231
unemployment (E)	33.500	23.961
E × J	2.821	2.018
E × S	0.142	1.422
Income (I)	1.273	0.911
I × J	6.772	4.843
I × S	0.331	3.319
C × E	0.025	0.018
C × E × J	0.049	0.035
C × E × S	0.001	0.014
C × I	0.044	0.031
C × I × J	0.000	0.000
C × I × S	0.004	0.044
E × I	0.111	0.079
E × I × J	0.244	0.174
E × I × S	0.021	0.215
C × E × I	0.000	0.000
C × E × I × J	0.014	0.010
C × E × I × S	0.001	0.006

Table of Contents

Table 70: MATH-CPS model results under reform 11 using 5 sources of uncertainty, and seed set A. A jackknife of 8 rotation groups was used with 2 seeds nested.

	full run	single run
$\hat{\beta}$	.291 (.113)	.291 (.422)
$\hat{\beta}_C$	-.047 (.083)	-.047 (.088)
$\hat{\beta}_E$	-.076 (.326)	-.076 (.331)
$\hat{\beta}_I$	.004 (.093)	.004 (.113)
$\hat{\beta}_{CE}$	.005 (.016)	.005 (.019)
$\hat{\beta}_{CI}$	-.001 (.011)	-.001 (.013)
$\hat{\beta}_{EI}$	.004 (.006)	.004 (.021)
$\hat{\beta}_{CEI}$	.000 (.003)	.000 (.003)
$\sqrt{\sigma_h^2/n}$	0.0	
$\sqrt{\sigma_s^2/n}$	0.422	
SE(true mean | estimates)	0.378	0.563
95% interval	(-0.465, 1.048)	(-0.835, 1.417)
% var due to:		
sampling (J)	0.000	0.000
seed (S)	8.888	56.145
calib (C)	1.549	0.699
C × J	4.791	2.162
C × S	0.047	0.299
unemployment (E)	4.019	1.814
E × J	74.165	33.463
E × S	0.170	1.073
Income (I)	0.010	0.004
I × J	5.821	2.627
I × S	0.221	1.397
C × E	0.016	0.007
C × E × J	0.172	0.078
C × E × S	0.006	0.036
C × I	0.000	0.000
C × I × J	0.085	0.038
C × I × S	0.002	0.013
E × I	0.011	0.005
E × I × J	0.000	0.000
E × I × S	0.022	0.139
C × E × I	0.000	0.000
C × E × I × J	0.005	0.002
C × E × I × S	0.000	0.001

Table of Contents

A.4 Overview of relevant MATH-CPS modules

DEMAGE - demographically ages MATH-CPS file

DEMAGE adjusts family and person-level weights "using multiplicative factors", so that population counts by age, race and sex agree with projected counts. The user can specify whether or not weights are to be adjusted. Also the user can specify the simulation year, and give a 16 x 8 array of population projections (16 ages by 8 race/sex categories). The output is family and person weights after aging.

ECONAGE - adjusts annual income variables for economic and population growth

Each type of wage is "aged", according to the educational class of the individual. Educational classes are less than HS, HS, more than HS. Also annual poverty threshold is adjusted.

ALLOY - allocates income to month

For people who worked part of the year, ALLOY first calculates the number of weeks worked. The start date for employment is assigned at random. The period of employment is assigned to start at that time, and continue until the number of weeks of employment is exhausted. Earnings are assigned uniformly across the period of employment.

Weeks for unemployment and not in labor force are then assigned. A random number determines whether unemployment or not in labor force follows employment (if the person was both unemployed and not in labor force during parts of the year).

ASSETS - imputes financial and vehicular assets

Six regression equations are used for imputing family financial and vehicular assets. The dependent variables modeled are: (1) family financial assets (the X's include family earnings, whether or not the family owns their house, and size, race, and education of the head, etc); (2) family car count (based on many of the same variables and others); (3) value of first family car; (4) ratio of equity to value for the first family car; (5) average value of remaining family cars; (6) average equity of remaining family cars. In each case a random error is added to the equation.

The household's assets affect eligibility for public assistance. About 50% of income-eligible FSP households have countable assets, averaging $12,000. 21% have vehicular assets averaging $4,000. Assets make 26% of the income-eligible households asset-ineligible for food stamps. Also, 17% of AFDC families have assets, with an average of $330 in total (financial plus vehicular) assets.

Calibration is used for financial and vehicular assets. For financial assets, 25% of households with income less than 130% of poverty and no assets are randomly selected to get $10 in asset income. The estimated amount of assets is reduced by 20%. For vehicles, "negative adjustment factors" are applied to the equation that estimates number of vehicles, and the value of the first vehicle is reduced by 20% whereas the value of subsequent vehicles is reduced by 10%.

CHILDEXP - imputes child care expenses to families

The first regression equation models whether or not the family has child care expenses. The second equation models the expense, conditional on the family having child care expenses.

The model for whether or not a family has child care expenses, models a "no" if there are no children under 15, and "no" if the guardian is not working. Otherwise, the family has a child care expense if $-\epsilon \leq X\beta +$ additive calibration factors, where the X's include sex, race, marital status, age category, and education of the guardian, number of people in the family, average age of children by category, log of earned and income, whether the household rents or not, number hours the guardian works per week, and log(guardian's hourly earnings).

The expenses, for family with expenses, are equal to min((wks per month [worked?] in previous year $\times \exp^{X\beta+\epsilon}$, 433) $\times$ multiplicative factors), adjusted to simulation year dollars. Here X's include guardian's presence, race, age category, education category, log(unearned income), whether the household rents or not, hourly earnings for guardian and for spouse, and guardian's hours worked per week.

Child care expenses affect eligibility for AFDC and FSP. About 5% of FSP households have a child care expense deduction, with an average deduction of $164.

Table of Contents

Both equations are calibrated. .5 is added to the first equation. Estimated expenses are calibrated to be 5% less.

UNIT7: Defines filing units for SSI, AFDC, and GA

This subroutine carries out rules to determine who is categorically eligible for public assistance.

PBLAST: Simulates means eligibility for SSI, AFDC, and GA

Means eligibility, simulated through PBLAST, depends on monthly income, computed through ALLOY. The user can affect this with parameters which govern the definition of countable income and assets, eligibility limits, maximum and minimum payments, etc. At the users option, some units are selected to be "new" to the AFDC program. The new units may be eligible for $30 and 1/3 earnings deductions. These units are selected using MTHRND.

PAPRAT: chooses units to participate in public assistance, from among eligible units

Participation in public assistance (PA) is simulated in a hierarchical fashion with the following precedence: AFDC-UP, AFDC-Basic, SSI-disabled, SSI-aged, and GA. Thus for example if there is an eligible AFDC unit in the household, then the AFDC participation probability is used, and if the household is simulated to participate in AFDC then the eligible SSI and GA units will also be simulated to participate.

Participation probabilities depend on region (usually CA, NY, TX, and other states - different for general assistance), type of public assistance, and whether or not the household reported receiving welfare in the last calendar year. Any household which reports receiving welfare and in which the target number of participants is less than the number of eligible units (for the particular PA program) is modeled to participate. Otherwise the household is modeled to participate if $U_i \leq \alpha$(region, PA type), where α [i.e. the participation probability] is determined iteratively during database development, and is modified during program development to meet control totals. Reforms use the same probabilities, unless otherwise specified by the user.

Table of Contents

According the the output tables, all SSI reporters, all AFDC eligible households and all GA eligible households are selected to participate.

MEDEXP: imputes out-of-pocket medical expenses

This imputes the amount of out-of-pocket medical expenses for households which have an elderly or disabled person, based on a tobit equation. The amounts are adjusted to simulation year dollars, and additive and multiplicative factors are applied to calibrate the total to the predicted amount for 1996.

The medical expense is min(max(0, $\exp^{X\beta+\epsilon} - 1$), 1000), adjusted to simulation year dollars, and then calibrated. The X's include log(1980 income), age, education and race categories of the household reference person, number of people in the household, whether or not the household gets PA, and others.

Medical expenses affect medical expense deductions, which affects net income. This affects about 3% of FSP households, with an average medical expense deduction of $96.

The medical expenses are calibrated by first multiplying by .5, and then subtracting $50, for all households.

SHLTREXP: Imputes shelter expenses

This imputes shelter expenses (including for fuel and utilities) for households, separately for homeowners, and for renters. The amounts are adjusted to simulation year dollars. Then additive and multiplicative factors are used to calibrate the [totals] to the 1996 predicted amount. Shelter expenses are not used for people living in group quarters.

For owners, the shelter expense is min(max(0, $\exp^{X\beta+\epsilon_1}$), 1212), and for renters the shelter expense is min(max(0, $\exp^{X\beta+\epsilon_2}$), 1036). The X's include log(monthly income), dummy variables for age, education, region, size of city, and race, household size, and whether or not the household receives welfare.

Shelter expenses affect shelter deductions, which affects net income. Among FSP households, 63% have excess shelter expense deductions, averaging $162.

The equations are calibrated by first multiplying shelter expenses by 1.2. Then $100 is subtracted from shelter expenses for rented households with income less than 50% of poverty, and $50 is subtracted from shelter expenses

Table of Contents

for rented households with income between 50 and 100% of poverty.

FSTAMP: simulates participation in the food stamp program

Any FSU in which all members are on PA are automatically eligible for food stamps. Otherwise, if the FSU passes the net and gross income tests and the asset test, and the computed FS benefit is positive, the FSU is eligible. Participation is simulated based on unit size, receipt of PA, presence of elderly, reporter status, and gross monthly income relative to poverty. At the user's option, the probability can also depend on whether or not the unit reports receiving food stamps. Separate probabilities are estimated for units receiving AFDC.

The sequence for coming up with the participation probabilities for households which do not receive AFDC is: (1) In a 4-way matrix indexed by income relative to poverty (4 classes), unit size (1,2,3-5, >5), whether or not SSI or GA is received, and presence or absence of elderly, the probabilities are the number of participants (based on summer 1991 IQCS control totals) divided by the number of eligibles (from the MATH-CPS model simulation). (2) Counts in cells in which the number of participants is greater than the number of eligibles are reallocated to other cells (results are fairly different from (1)). (3) The participation matrix is expanded to include the benefit amount / poverty, and at the user's option to include reporter status.

These probabilities are 1 for everyone for whom the poverty ratio (the ratio of income to poverty) is between 0 and .50, and for many units without elderly at all levels of the poverty ratio. The probabilities are 0 for many units with elderly and for whom the poverty ratio is greater than 1.

These probabilities, for baselaw participation, may be (and in our program, are) further adjusted (in PPART) by multiplying them by has_earn_factor, where has_earn_factor = .50 if the unit has earners and receives SSI, has_earn_factor = .75 if the unit has earners and does not receive SSI, has_earn_factor = .95 if the unit has no earners and receives SSI, and has_earn_factor = 2.00 if the unit has no earners and does not receive SSI.

For households which receive AFDC, the participation probability equals a user defined participation rate × number participants (in FSP?) ÷ number eligible for AFDC, where the number of eligibles is from the number of AFDC eligibles counted in the MATH-CPS model. For the baselaw, the [default] user defined participation rate is .3960, and total number of participants is 10,864,000. [I think this means 39.6% of the 10,864,000 FS participants are

Table of Contents

on AFDC]. According to their tables, all households which participate in AFDC also participate in the FSP.

The participation probabilities for the reform depend on eligibility in the reform, and eligibility and participation in the baselaw. There are 5 cases: (1) The unit was eligible in the baselaw but didn't participate, and unit would receive a larger benefit under the reform than the baselaw. In this case the probability of participation, mu, is determined by the refalgo subroutine. (2) The unit didn't participate in the baselaw and receives a smaller or the same benefit under the reform. In this case, force the unit not to participate. (3) The unit participated in the baselaw and would receive the same or larger benefit in the reform. Force the unit to participate. (4) The unit participated in the baselaw and would receive a smaller benefit in the reform. The participation probability is 1 - mu, where mu is determined from refalgo. (5) The unit is not eligible in the baselaw. In this case, use the baselaw method to calculate the participation probability.

For (1) and (4) above, mu is calculated from the refalgo subroutine. Mu equals mills $\times \beta_0 \times$ |(benefit in reform - benefit in baselaw)| $\div$ benefit in baselaw. If the unit is better off in the reform, mills $= \frac{pdf(X\beta)}{1-cdf(X\beta)}$. Otherwise (unit does the same or worse in reform), mills $= \frac{pdf(x\beta)}{cdf(X\beta)}$.

Here X's are dummy variables for household size, ratio of household income to poverty level, age, race and education of reference person, number of children.

Table of Contents

A.5 Comparison of methods for determining FSP participation probabilities

Our algorithm for allocation excess participants in the 64-cell matrix used for determining food stamp program participation rates, differs from the current MPR method. Our algorithm is based on an algorithm originally considered by MPR, coded in a program called fsprob96.f. The algorithm used by fsprob96.f to reallocate participants was originally thought by MPR to be sufficient, but was found not to be.

Since our method of reallocating participants is based on the fsprob96.f algorithm, we first describe the iterative fsprob96.f algorithm. First the total number of eligibles and the total number of extra participants are calculated. Extra participants are the excess number of participants in cells for which the number of participants is greater than the number of eligibles. The ratio of the total number of extra participants to the total number of eligibles is then added to the probability in any cell for which the probability is less than 1, and the number of participants is adjusted accordingly. Probabilities in cells in which the number of participants is greater than the number of eligibles are adjusted to be 1.0, and the number of participants in these cells are set to be the number of eligibles. The method is iterative, in that after this adjustment is made to each cell, the loop continues to execute until no cell has a probability greater than 1.

Our method is based on the ideas in fsprob96.f. However, we only reallocate excess participants in a given cell to adjacent cells, where adjacent cells are defined to be any cells which share 3 out of the 4 dimensions with the given cell. Another difference is that we add excess participants to any adjacent cell, regardless of whether or not the adjacent cell probability is greater than one. This results in the same end probabilities regardless of the order in which cells are processed.

Below we show the initial number of eligible households, initial numbers of participants and initial probabilities. We also compare numerical results from the MPR method and our method.

Table of Contents

ELIGIBLES (Source: MATH-CPS model)

	Unit Size 1	2	3-5	6+
Poverty Ratio < .50				
No PA, no elderly	517319.	207560.	377705.	69709.
No PA, w/ elderly	308037.	124590.	15988.	4450.
w/ PA, no elderly	196143.	30501.	85656.	11987.
w/ PA, w/ elderly	9698.	4052.	7233.	5799.
.50 < Poverty Ratio <= 1.00				
No PA, no elderly	286646.	255300.	837588.	135271.
No PA, w/ elderly	1738114.	426908.	70600.	14005.
w/ PA, no elderly	669299.	225383.	70134.	8185.
w/ PA, w/ elderly	914773.	224997.	32453.	12666.
1.00 < Poverty Ratio <= 1.30				
No PA, no elderly	283822.	278688.	584647.	83000.
No PA, w/ elderly	1005777.	387444.	53009.	5685.
w/ PA, no elderly	27281.	78239.	101048.	6800.
w/ PA, w/ elderly	12682.	127052.	29805.	0.
Poverty Ratio > 1.30				
No PA, no elderly	11329.	17107.	5109.	1136.
No PA, w/ elderly	196286.	129504.	10043.	1271.
w/ PA, no elderly	7994.	39640.	16084.	0.
w/ PA, w/ elderly	3235.	56357.	2568.	0.

Table of Contents

NUMBER OF PARTICIPANTS (Source: summer 1992 IQCS)

794195.	293312.	384491.	53917.
44438.	19703.	5639.	0.
285308.	37748.	56890.	3657.
23838.	8006.	13585.	1154.
299528.	275845.	739920.	67765.
266888.	56813.	18200.	1679.
712636.	95337.	57082.	14485.
682203.	108651.	20797.	1136.
82182.	111998.	216912.	26297.
121539.	35836.	9120.	0.
22224.	11138.	14123.	1805.
19598.	17463.	6780.	0.
7750.	4098.	1579.	348.
17378.	2636.	0.	0.
3876.	1564.	0.	0.
2405.	348.	0.	0.

Initial probabilities

1.5352	1.4131	1.0180	0.7735
0.1443	0.1581	0.3527	0.0000
1.4546	1.2376	0.6642	0.3051
2.4580	1.9758	1.8782	0.1990
1.0449	1.0805	0.8834	0.5010
0.1536	0.1331	0.2578	0.1199
1.0647	0.4230	0.8139	1.7697
0.7458	0.4829	0.6408	0.0897
0.2896	0.4019	0.3710	0.3168
0.1208	0.0925	0.1720	0.0000
0.8146	0.1424	0.1398	0.2654
1.5453	0.1374	0.2275	0.0000
0.6841	0.2396	0.3091	0.3063
0.0885	0.0204	0.0000	0.0000
0.4849	0.0395	0.0000	0.0000
0.7434	0.0062	0.0000	0.0000

Table of Contents

Probabilities from MPR algorithm

1.0000	1.0000	1.0000	1.0000
1.0000	1.0000	1.0000	1.0000
1.0000	1.0000	1.0000	1.0000
1.0000	1.0000	1.0000	1.0000
0.9855	0.9950	0.9069	0.4270
0.1651	0.0524	0.1161	0.1199
0.9971	1.0000	0.8244	1.0000
0.7520	0.2607	0.1786	0.0897
0.9868	0.2402	0.3796	0.3504
0.0314	0.0000	0.0000	0.0000
0.9613	0.6408	0.4762	0.9509
0.2115	0.0000	0.0000	0.0000
1.0000	0.2396	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000
1.0000	0.8972	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000

Probabilities after using Sally Thurston's algorithm

1.0000	1.0000	1.0000	1.0000
0.4062	0.2613	0.4484	0.0094
1.0000	1.0000	0.8941	0.4549
1.0000	1.0000	1.0000	0.2518
1.0000	1.0000	0.9914	0.5637
0.1793	0.1545	0.2578	0.1199
1.0000	0.5498	0.9029	1.0000
0.8427	0.4901	0.6757	0.0959
0.5668	0.5192	0.4319	0.3263
0.1242	0.0925	0.1720	0.0000
1.0000	0.1592	0.1401	0.2720
1.0000	0.1480	0.2657	0.0000
0.9610	0.3569	0.3699	0.3158
0.0885	0.0204	0.0000	0.0000
0.6858	0.0559	0.0000	0.0000
0.7576	0.0133	0.0348	0.0000