REWEIGHTING HOUSEHOLDS TO DEVELOP MICROSIMULATION ESTIMATES FOR STATES

Allen L. Schirm, Mathematica Policy Research, Inc. and Alan M. Zaslavsky, Harvard University
Allen Schirm, Mathematica Policy Research, 600 Maryland Ave., S.W., Suite 550, Washington, DC 20024-2512

Key Words: Small Area, Welfare Reform, Food Stamps

1. INTRODUCTION

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) made sweeping changes to our welfare system, including giving states great flexibility to design their own programs. States have begun to exercise this new authority in order to satisfy broad requirements of PRWORA and qualify for funding under the Temporary Assistance to Needy Families (TANF) block grant established by the act. However, even several months after the July 1, 1997 deadline for submitting TANF plans to the federal government, some states are operating essentially placeholder programs that they expect to change substantially during the next year. Nearly all states will continue to refine their programs for years to come. As state policymakers consider both large and small program reforms, they will need reliable estimates of the cost and distributional effects of alternative policies.

Microsimulation has been used extensively to obtain national estimates for understanding the effects of proposed reforms to federal programs. However, the current microsimulation models that use data from national surveys can produce only highly imprecise estimates for states because sample sizes are small.

This paper describes a method for reweighting a microsimulation database to "borrow strength" and obtain more precise state estimates. A Poisson regression model is fitted to obtain an estimated prevalence in each state of every household in the database. This model is specified to control important aggregates at the state level, and the prevalences are expressed as a matrix of weights, with each household having a weight for every state. Estimates for a state are obtained by passing through the microsimulation model all households in the database, not just the households actually observed in that state. By applying the appropriate weight for each household, the database is weighted to look like the state, rather than the whole country. We describe this approach and a preliminary evaluation of it in this paper after briefly reviewing how microsimulation methods are used for policy analysis.

2. WHAT IS MICROSIMULATION?

A microsimulation model simulates how proposed changes to a government program affect the program and program participants. The model has two elements: (1) a micro database and (2) a computer program. The database is constructed from administrative or survey data with information on households in the population targeted by the government program. The model's computer program codes the rules of the government program under both the "baseline" policy, which is typically the current policy, and a "reform" policy, which is an alternative under consideration. The computer program also simulates what a caseworker does--that is, it determines whether a household is eligible for the government program and the benefits for which the household would qualify. In addition, the computer program simulates a household's behavioral response, determining whether the household will participate in the program. Processing all the households in the database, the model counts participants to estimate the caseload of the government program and adds up their benefits to estimate costs. By performing these operations under both baseline and reform policies and comparing the results, the model estimates the cost and caseload effects of the proposed reform. The model can also estimate the distributional effects of the reform, identifying the population subgroups that gain and lose benefits.

From among the microsimulation models currently in use, we focus in this paper on the MATH SIPP model, the Micro Analysis of Transfers to Households (MATH®) model that uses data from the Survey of Income and Program Participation (SIPP). The MATH family of models, developed by Mathematica Policy Research beginning in 1974, has been used extensively to simulate reforms to the Food Stamp Program (FSP), the Aid to Families with Dependent Children (AFDC) and TANF programs, and the Supplemental Security Income (SSI) program. The database for the current MATH SIPP model was constructed by combining data for January 1994 from Waves 7 and 4 of the 1992 and 1993 Panels.

3. MICROSIMULATION FOR STATES

Although the MATH SIPP model has been used in recent years to estimate the national effects of hundreds of potential reforms to national programs (mainly the FSP), welfare devolution and the resulting need for state analyses and estimates create a new challenge. Our main statistical problem is that because a database developed from a national survey like the SIPP or even a national administrative database will have small sample sizes for most states, direct microsimulation estimates will be imprecise. Thus, there will be substantial uncertainty about the likely impacts of a proposed program reform.

Borrowing strength with an indirect estimator is a general solution to the problem of imprecise direct estimates, and has been used successfully in many applications. For example, an indirect estimator has been used for several years to obtain state estimates for allocating federal funds under the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). Estimates of the numbers of infants (under age 1) and children (ages 1 to 4) below 185 percent of poverty are derived in three steps: (1) calculate direct sample estimates from the Current Population Survey (CPS); (2) calculate regression estimates, that is, predictions based on census and administrative data (measuring, for example, food stamp and unemployment insurance program participation); and (3) calculate "shrinkage" estimates by averaging the sample and regression estimates using empirical Bayes methods. The shrinkage estimates are substantially more precise than the direct sample estimates.

In the microsimulation context, the problem with using an empirical Bayes estimator like the WIC eligibles estimator or a similar estimator that smooths noisy sample estimates is that such methods are not "off-the-shelf" approaches for a microsimulation model's many estimands, which include baseline and reform caseload and cost figures as well as numerous measures of distributional effects. Instead, these "standard" indirect estimators are best for deriving a single estimate or a few closely related estimates for each state. A different approach to indirect estimation is needed for microsimulation. We propose to borrow strength and improve precision by reweighting sample observations in the microsimulation database.

4. REWEIGHTING TO BORROW STRENGTH AND IMPROVE PRECISION

How reweighting can be used to borrow strength is illustrated by comparing (1) the direct estimator that uses the original sample weights and does not borrow strength with (2) the indirect estimator that uses reweighted data and does borrow strength. To calculate an estimate for Virginia, for example, the direct estimator uses only the sample households for Virginia and their original sample weights. Observations for other states are ignored. This is equivalent to using all the observations in the database weighted by "Virginia weights" that equal the original sample weights for households in Virginia but are zero for households in all other states. In contrast, for indirect estimation, nonzero Virginia weights would be assigned to households in not only Virginia but also other states. Thus, the indirect estimator uses households outside Virginia to derive estimates for Virginia, thereby borrowing strength.

We propose to assign weights for indirect estimation using a method developed by Zaslavsky (1990). With this method, we derive a matrix of weights. Every household in the database gets as many new weights as there are states. For every state, there is a weight for each household in the database, although some weights may turn out to be zero or nearly zero. Virginia weights are used to derive estimates for Virginia, North Carolina weights are used to derive estimates for North Carolina, and so forth.

A household is assigned weights depending on its "type," which is defined by all the characteristics in the database, some of which are measured directly in the survey while others are calculated or simulated. Each sample household is regarded as representative of households of the same (or similar) type in its home state and all other states.

Using a Poisson regression model described later, our reweighting method assigns a Virginia weight to a household according to how prevalent that type of household is in Virginia. The more prevalent it is, the more Virginia weight it gets. Prevalence is determined, under the model, by a set of household characteristics that (1) capture the key dimensions along which households in different states are different and (2) are policy-relevant (e.g., characteristics determining program eligibility and benefits).

As will become clearer later when we present the formal reweighting model, these characteristics variables serve as control variables, and households are reweighted so that weighted sums (indirect estimates) equal specified control totals. These totals can be direct sample estimates, indirect estimates smoothed using empirical Bayes or other methods, or administrative totals. For example, if the number of people in the household is a control variable, the total state population is a control total. If household income is a control variable, total personal income in the state is a control total.

Before reweighting, the database is weighted to look like the entire United States. After reweighting and, specifically, after the derivation of Virginia weights, the database is weighted to look like Virginia in terms of some key aggregates (the control totals). Then, we hope that the reweighted database resembles Virginia in terms of many other relevant aggregates for which we cannot control, including, especially, the main estimands of the microsimulation model. The extent to which this is accomplished is an important criterion for assessing the method's success.

Our reweighting method redistributes weight from Virginia households to households in other states, which may not seem like a sensible strategy. However, we have to take some weight away from Virginia households so that we do not change Virginia's population, that is, the weighted number of Virginians calculated across all observations in the database. Although redistributing weight from Virginia to non-Virginia households may introduce some bias because the non-Virginia households are not exactly like Virginia households, the redistribution should substantially improve precision by allowing us to use many more observations to obtain estimates for Virginia. This will be especially important for policy-relevant subgroups that have only a few observations in the Virginia sample. The redistribution of weight, moreover, is systematic because households are weighted according to how prevalent they would be if they were in Virginia. Giving some weight to non-Virginia households that are similar to Virginia households except for state of residence should improve precision without introducing substantial bias. The objective of reweighting and, more generally, indirect estimation is to enhance the accuracy of estimates as measured by a standard like mean squared error (MSE) that reflects the tradeoff between bias and variance.

As noted earlier, the control totals used in reweighting households can be obtained from different sources and chosen to satisfy different objectives. For example, weight can be redistributed to replicate features of the Virginia population estimated using the original sample weights. If the direct sample estimate of the number of household members receiving Medicaid is a control variable, we will reweight observations so that the direct and indirect estimators give the same number of Medicaid recipients in Virginia, even though the indirect estimator gives positive weight to households outside Virginia. Alternatively, we may choose to reweight observations to adjust population quantities estimated from the sample so that indirect estimates agree with administrative totals, rather than direct estimates. (Such adjustment, when it is not also used to borrow strength, is called "calibration.") To correct for undercoverage of Medicaid recipients and underreporting of Medicaid recipiency, we could reweight observations so that the indirect estimate of Medicaid recipients in Virginia equals the total number of recipients in Virginia's Medicaid case records. We note in this regard that if the database has been calibrated prior to reweighting for indirect estimation, the reweighting can be done so as to preserve the calibration.

5. A PRELIMINARY REWEIGHTING

We have completed a preliminary reweighting of the MATH SIPP database according to the first three steps of the following procedure:

1. *Identify control variables.* The variables should capture important differences among households in different states and be relevant to current programs or program reforms that may be proposed. All control variables have to be directly measured in or calculable (perhaps simulated) from the microsimulation database. The same set of control variables is used for each state.

2. *Obtain control totals.* These can be (i) direct sample estimates, (ii) indirect (e.g., shrinkage) estimates, or (iii) administrative totals.

3. *Fit Poisson regression model and calculate weights.* The model is estimated so that (i) the total weight given to a household (in constructing a national estimate) is not changed and (ii) all control totals are satisfied for every state.

4. *Assess model fit, iterating back to (1) as necessary to improve fit.*

5. *Produce final weights.*

After completing our preliminary reweighting of the MATH SIPP database, we have begun assessing model fit. Before discussing our early findings, we will elaborate on our implementation of the first three steps.

For the preliminary reweighting, we divided households in the database into three income groups: low (at or below 130 percent of poverty), medium (130 to 300 percent of poverty), and high (above 300 percent of poverty). We chose these groupings because the FSP gross income threshold is set at 130 percent of poverty, and very few households above 300 percent of poverty are eligible for any means-tested transfer program. We fit a model for each group with the following control variables (all measured at the household level): an intercept, number of persons, number of Hispanics, number of blacks, number of Asians, presence of members unrelated to the head (0/1), presence of kids under age 5 (0/1), presence of an elderly member (0/1), shelter expenses, utility expenses, home ownership (0/1), receipt of interest income (0/1), receipt of earned income (0/1), receipt of SSI income (0/1), number of working age adults (ages 18 to 59), number of working age adults with less than a high school education, and number of working age adults with no job. These variables, which measure household composition, income, and expenses, were selected based on their policy relevance and analysis of variance results showing that their distributions varied across states. In all instances, the control totals used for the preliminary reweighting are direct sample estimates.

The Poisson regression model used for reweighting is:

$$\ln w_{hs} = \beta_s' x_h + \delta_h,$$

where w_{hs} is the expected number of households of type h in (the population of) state s. A type is a complete household record and is, practically speaking, unique on the database because no two households are exactly alike. Therefore, each household in our database represents its own type, and w_{hs} is the weight that will be given to household h when deriving estimates for state s. x_h is a column vector of I control variables: $x_h = (x_{h1}\ x_{h2} \dots x_{hI})'$. It gives the values of those variables for household h. β_s is a column vector of I unknown parameters--the regression coefficients for the control variables--to be estimated for each state. δ_h is an unknown parameter to be estimated for each household. The first term on the right side of the Poisson regression model ($\beta_s' x_h$) reflects the "general" prevalence in state s of households like household h, that is, households with similar characteristics. The second term (δ_h) reflects the "specific" prevalence of household h.

The β_s and δ_h parameters are estimated by maximum likelihood and satisfy two constraints:

$$\text{Constraint 1: } \sum_s w_{hs} = W_h \quad \text{for each } h,$$

where W_h is the control weight, that is, the original sample weight of household h, and

$$\text{Constraint 2: } \sum_h w_{hs} x_{hi} = X_{si} \quad \text{for each } s \text{ and } i,$$

where X_{si} is the control total for control variable i in state s. According to the first constraint, reweighting does not change the total weight given to a household across all states, that is, at the national level. This constraint ensures that a household contributes the same to a national estimate after reweighting as before. Reweighting does change a household's contributions to estimates for individual states, enabling strength to be borrowed and precision to be improved. According to the second constraint, all control totals are satisfied for every state. That is, the weighted sum of a control variable after reweighting equals the specified control total.

We estimated the Poisson regression model using an iterative two-step procedure. If $\beta_{s(k)}$ and $\delta_{h(k)}$ are the values for the unknown parameters in iteration k, the two steps are:

$$\text{Step 1: } \delta_{h(k)} = \ln\left(\frac{W_h}{\sum_s e^{\beta_{s(k-1)}' x_h}}\right)$$

and

$$\text{Step 2: } \beta_{s(k)} = \beta_{s(k-1)} + D_s^{-1} d_s \quad \text{for each } s,$$

where

$$D_s = \sum_h w_{hs} x_h x_h',$$

$$d_s = X_s - \sum_h w_{hs} x_h,$$

and w_{hs} is obtained by substituting $\beta_{s(k-1)}$ and $\delta_{h(k)}$ into the expression for the Poisson model. X_s is a column vector of control totals for state s, with one control total for each of the I control variables: $X_s = (X_{s1}\ X_{s2} \dots X_{sI})'$. d_s is a column vector (with I elements) giving the differences between weighted sums and control totals at the beginning of the second step of iteration k. When each element of d_s is sufficiently small according to a prespecified convergence criterion (e.g., $d_{si} < 0.01$), the weights from the Poisson model approximately satisfy the second constraint. If the second constraint is satisfied before the second step, that step need not be completed, and because the first constraint was satisfied at the end of the first step, no further iterations are required.

As this description suggests, each of the two steps of the proposed estimation procedure is designed to satisfy one of the two constraints. The first step of the estimation procedure is obtained by substituting the expression for w_{hs} implied by the Poisson model into the first constraint and solving for δ_h. Thus, carrying out the first step satisfies the first constraint. The second step of the estimation procedure is a single Newton-Raphson step toward satisfying the second constraint. At this time, we do not recommend performing additional Newton-Raphson steps until convergence is achieved, that is, until the second constraint is fully satisfied in iteration k. The estimation procedure should be more efficient if just a single Newton-Raphson step is performed in each iteration because the convergence of the Newton-Raphson step is quadratic and, therefore, faster (when near convergence) than the convergence of the overall algorithm, which is linear. After "enough" iterations of the two-step procedure (the number of iterations depending on the convergence criterion), both constraints will be satisfied.

6. PRELIMINARY EVALUATION RESULTS

We have begun our evaluation of the preliminary reweighting of the MATH SIPP database by comparing indirect (model-based) estimates derived using the preliminary state weights to direct sample estimates derived using the original sample weights. In this paper, we report comparisons for two key estimands from our microsimulation model. The first is total food stamp benefits under baseline (current) policy. Because total benefits vary widely by state size, we have divided total benefits by the total number of persons, one of our control variables, to obtain per capita total benefits. The second estimand is the impact on total benefits of a fairly generic program reform, which raises the food stamp earnings deduction from 20 to 30 percent. Comparisons of model and sample estimates for these two estimands should be a good first test of our approach because although we have controlled for the total population size of each state, the number of households with earnings in each state, and other aggregates, we have not controlled directly for total food stamp benefits.

In Figures 1a and 1b, we have graphed the model estimate against the sample estimate for each of the 42 states uniquely identified in SIPP public use files. As expected, model and sample estimates are different. However, there does not seem to be a systematic pattern to the differences except for some regression to the mean, which has two indistinguishable sources--one "good" and one "bad." The bad one is model error, that is, the failure to capture true differences among states, and, specifically, the tendency for a model to overpredict at the low end and underpredict at the high end. The good one is the smoothing away of noise in the sample estimates. This corrects for the tendency of some high sample estimates to be high because of large positive sampling errors and some low sample estimates to be low because of large negative sampling errors. The pulling in of extreme values is most apparent for reform impact estimates (Figure 1b).

Although some differences between model and sample point estimates are expected and, as suggested, desired, the foremost purpose of reweighting is to improve precision. By that standard, the preliminary reweighting has succeeded according to Figures 2a and 2b, where we have plotted estimated standard errors for model estimates against estimated standard errors for sample estimates. To estimate standard errors, we split the database into 40 random groups (after sorting by state and SIPP pseudo-sampling stratum within state) and used a grouped jackknife estimator, treating weights as fixed.

Two findings from our comparison of standard errors are most striking. First, the variability of sample estimates is unacceptably large. Second, the variability of model estimates is substantially less and at a level at which the estimates can provide useful guidance to policymakers. The median coefficient of variation (CV)--the ratio of the standard error to the point estimate--is 34 percent for sample reform impact estimates, whereas the median CV is just 6 percent for model estimates. Although the CVs of sample estimates exceed 20 percent for all but 6 states, the CVs of model estimates are under 10 percent for all but 4 states. We find broadly similar patterns and gains from modeling for per capita benefits. That variances of model estimates are at least 90 percent smaller than variances of sample estimates for all but a few states (and for both estimands) suggests that unless the biases of the model estimates are enormous, the model estimates will have smaller MSEs than the sample estimates.

As a first step in evaluating the preliminary reweighting of the MATH SIPP database, we have compared in Figures 1a and 1b our model estimates with sample estimates, using the sample estimates as a measure of truth. However, the reason for reweighting the database is that the sample estimates are a very noisy measure of truth, as confirmed in Figures 2a and 2b. A natural question is whether the differences between model and sample estimates observed in Figures 1a and 1b are consistent with the sampling variability depicted in Figures 2a and 2b.

In Figures 3a and 3b, we have plotted differences between model and sample estimates (measured along the vertical axis), and displayed estimated two standard error bars, which give 95 percent confidence intervals for the differences. States are ranked (1 to 42) along the horizontal axis by their sample estimates. According to these two graphs, the estimated bars cross the horizontal line at zero for all but a small number of states, suggesting that the differences between model and sample estimates can generally be attributed to noise--mainly noise in the sample estimates. We should note that four of the seemingly significant differences pertain to states with sample reform impact estimates equal to zero because no sample household was affected by the reform. For these states, our naive jackknife estimator estimates that the standard error of the sample estimate is zero because there is no variability among sample households. That, of course, is not a sensible estimate.

Taken together, the graphs presented in this paper show that our reweighting approach succeeds in pulling in from either extreme the most extreme sample estimates, a typical phenomenon associated with indirect estimation and a desirable one. Even with this systematic pattern of smoothing, the estimated confidence intervals reveal that the "revised," that is, model estimates are (usually) fully consistent with what the sample estimates say, but are substantially more precise.

We are encouraged by these results, and will continue the evaluation by examining additional estimands of the microsimulation model. Recognizing the usual limitations of evaluating a model against the data used to fit the model, we will also apply cross-validation methods to assess predictive accuracy (as opposed to predictive fit). If we identify deficiencies of our reweighting models at any point in our evaluation, we will respecify the models. We also anticipate using administrative or indirect estimates as control totals rather than relying exclusively on direct estimates. We expect that would even further improve the accuracy of state microsimulation estimates.

REFERENCE

Zaslavsky, Alan. "Representing the Undercount by Multiple Imputation of Households." Technical report presented at the American Statistical Association Winter Conference. Cambridge, MA: Harvard University, Department of Statistics, January 1990.

Figure 1a. Model and Sample Estimates of Per Capita Food Stamp Benefits [Total Benefits/Total Population]

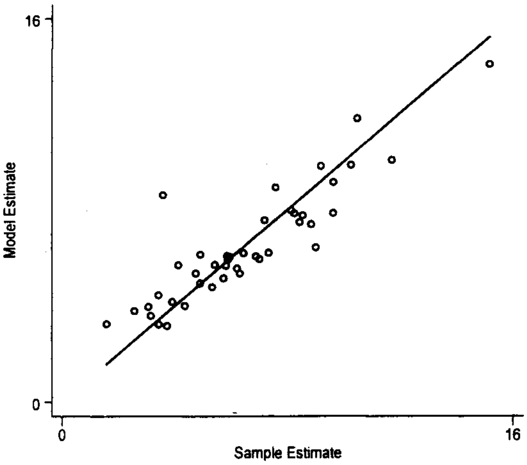

Figure 1b. Model and Sample Estimates of Reform Impact on Total Benefits [Percentage Change in Total Benefits]

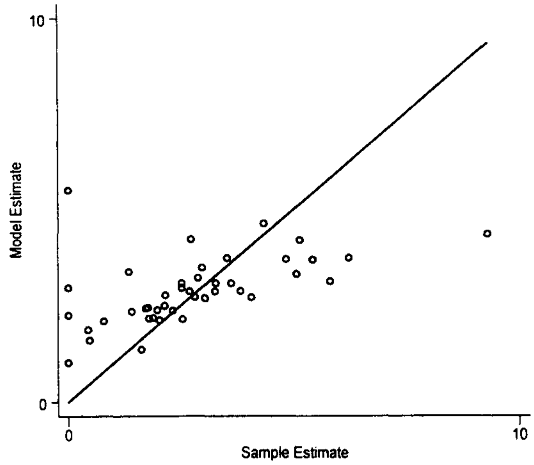

Figure 2a. Estimated Standard Errors of Model and Sample Per Capita Benefits Estimates

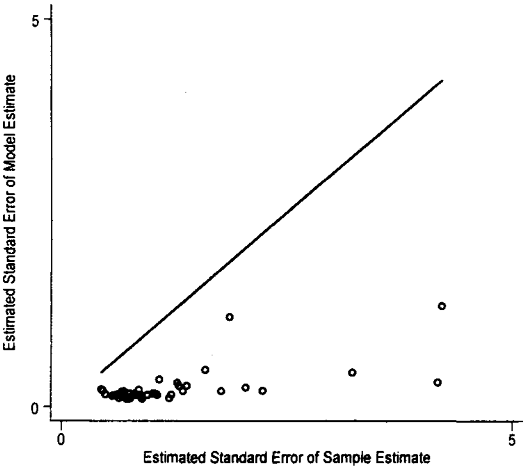

Figure 2b. Estimated Standard Errors of Model and Sample Reform Impact Estimates

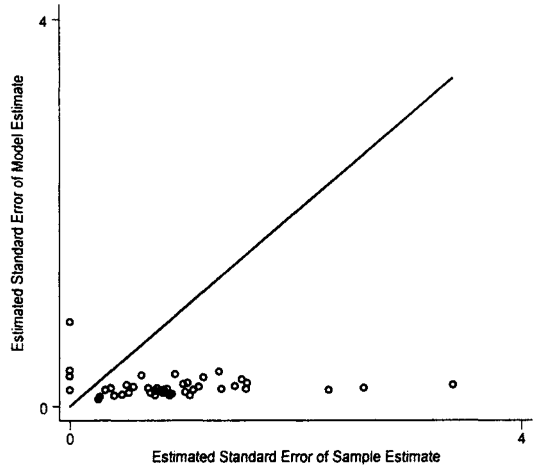

Figure 3a. Estimated Confidence Intervals for Differences Between Model and Sample Per Capita Benefits Estimates

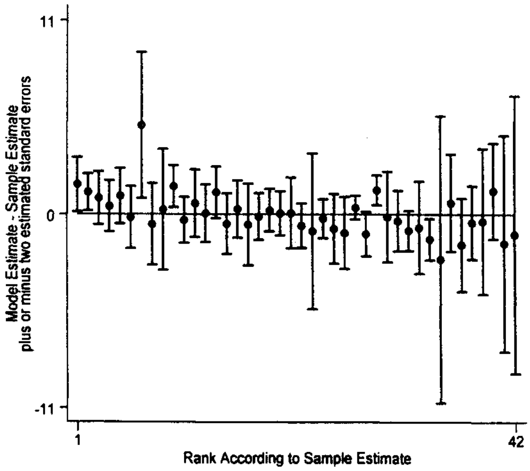

Figure 3b. Estimated Confidence Intervals for Differences Between Model and Sample Reform Impact Estimates