A PROPOSED ESTIMATION PROCEDURE FOR CPS GROSS FLOWS

Edwin L. Robison and Martha Duff, Bureau of Labor Statistics
Edwin Robison, BLS-OEUS-SMD Room 4985, 2 Massachusetts Avenue NE, Washington, DC 20212

Key Words: Gross Flows, Weighting, Raking, Current Population Survey

Abstract
Gross flows estimates describe the month-to-month transitions from one labor force state to another. The method currently in use for Current Population Survey (CPS) data applies a simple ratio adjustment forcing the weighted sum of matched cases (same person in sample two consecutive months) to duplicate overall male and female population control totals for the current month. The current month margin of a gross flows table does not duplicate the current month "stock" estimates of employed, unemployed, and not-in-labor force. An alternate gross flows weighting procedure for labor force estimates is proposed. Control totals are created for consecutive months by gender/labor force state by summing the full-sample CPS micro data weights for each month; adjustments account for persons flowing in-scope and out-of-scope between months. Matched cases are weighted, then raked and reweighed, to produce estimates that duplicate controls for both months.

Introduction

The CPS is a monthly rotating panel survey of about 60,000 households. Each month, labor force information is obtained from about 112,000 persons in the adult (16 years of age and older) civilian non-institutional population (CNP). In any given month, 8 panels of about 7,500 households are surveyed. The panels are sometimes labeled as MIS1 through MIS8. (Month-in-sample or MIS 1 through 8, the number-of-months a panel was included in the CPS). The panel rotation scheme is designed so that 6 of the 8 panels continue into the following month. CPS stock estimates of the labor force for a given month are made using weighted responses from all 8 panels. A monthly weighting process ensures that estimates of population made from the CPS duplicate a set of monthly population estimates produced by the Bureau of the Census.

Gross flows estimates, on the other hand, are restricted to the 6 panels in months-in-sample that continue from one month to the next. These are MIS1-MIS3 and MIS5-MIS7 in the prior month that become MIS2-MIS4 and MIS6-MIS8 in the current month. To estimate the total month-to-month changes from one labor force state to another, a weighting procedure must

be used to inflate the data obtained from the continuing portion of the sample (6 panels) up to the full complement of 8 panels.

Present Gross Flows Weighting Procedure

The present gross flows weighting procedure modifies the current month CPS final weights (CPSFW). In-scope responses all have a labor force classification of E (employed), U (unemployed), or N (not-in-labor-force). Using all 8 panels for the current month, the sum by gender of CPS final weights (CPSFW) duplicates the population control totals for men and women, CNP_m and CNP_w. Due to the rotation scheme of the CPS sample, only 6 of 8 panels are in common from one month to the next. While roughly 75% of the CPS sample is in continuing panels, matched cases represent only about 73% of the total sample. Final weights of matched persons in the 6 continuing panels are summed by gender (M_m for Men and M_w for Women). Gross Flows weights (GFW) are created by adjusting the CPS final weights for matched cases. These adjustments force the sums of matched cases to duplicate overall male and female control totals for the current month.

$$GFW_{mi} = CPSFW_{mi} * CNP_m/M_m \text{ for Men}$$
$$GFW_{wi} = CPSFW_{wi} * CNP_w/M_w \text{ for Women}$$

Gross Flows are then tabulated based on these adjusted weights on matched cases. Persons are classified into 3 labor force states and a 3 by 3 matrix is created to examine the prior-to-current month transitions (Table 1). In the table EU represents the number of persons who were employed in the prior month and are unemployed in the current month, etc.

Table 1. Gross Flows for Three Labor Force States		Current Month		
		Labor Force State		
		E	U	N
Prior Month	Labor Force State			
	E	EE	EU	EN
	U	UE	UU	UN
	N	NE	NU	NN

With this simple ratio adjustment to continuing panels, neither month's marginal E, U, N in Table 1 will duplicate the CPS stock estimates of E, U, and N. The difference between the prior month margin and the stock estimates is generally worse than for the current month, since only current weights are used in the

weighting. In addition, sub-tables, except those for male and female totals, will not add up to known population controls for either month.

Gross Flows Problems

Several general problems encountered in producing gross flows estimates are discussed briefly in this section.

Non-matched persons in matched panels: Some respondents in the panels continuing from one month to the next are not usable in gross flows estimation. Matched civilian labor force data for the same persons are needed, and not all persons are matched for various reasons that include:

- A household nonresponse in either month
- A person nonresponse for either month within a responding household
- Prior month persons who have since moved out
- Current month persons who have just moved in
- Persons who were age 15 in the prior month and are age 16 in the current month
- Prior month persons changing to out-of-scope in the current month
- Out-of-scope persons in the prior month joining the civilian labor force in the current month

Reporting errors: Errors in classifying the respondent into the correct labor force state have potentially significant effects on gross flow calculations. In stock data, classification errors tend to offset each other, while in flow data, errors tend to be additive.

Rotation group bias: Respondents may respond to labor force questions differently depending on how long they have been in the survey. It is known that the labor force state that a respondent reports in the CPS is affected by month-in-sample. For example, in 1996, the average (weighted) percentage of respondents not in the labor force ranged from 32.2 percent in MIS 1 to 33.2 percent in MIS 4 to 33.6 percent in MIS 8, with corresponding decreases in both percentage employed and percentage unemployed. This fact has clear implications for gross flows. Since MIS 1 and 5 cannot be matched to the prior month, the increase in not-in-the-labor-force for later MIS implies that matched samples will show flows out of the labor force even if there is no change in the stocks. This problem is observed in Gross Flows tables produced using the current method.

Inconsistent sets of weights: Prior-to-current month matched CPS persons have two sets of final weights—prior month final weights and current month final weights. Final weights for each month are our best attempt at adjusting for nonresponse and coverage errors, and producing approximately unbiased labor force estimates for that month. For a matched case, the prior month and current month final weights are slightly different due mostly to response rate differences, a linear growth in the population controls used in second-stage ratio adjustment, and interactions between persons induced by the second-stage ratio adjustment procedures.

One weight per matched person is needed for traditional gross flows analysis. The single weight is attached to the person's prior month status and also the person's current month status. Input-output matrices, like the one shown in Table 1, are then simple to construct.

Proposed Gross Flows Estimation Procedure

The proposed new weighting methodology directly addresses additivity problems in gross flows estimation due to non-matched cases, rotation group bias, and inconsistent weights. Our basic approach is to rake the flows in the matched CPS data to the stocks in each month's data, while taking into account flows into and out of the scope of CPS.

As a basis for the discussion of our method, Table 2 shows an expansion of the 3 by 3 gross flows table to deal with flows in and out of scope:

Table 2. Basic Gross Flows Table Structure.			Current Month					
			Labor Force State			Out-Flow		Row Total
			E	U	N	Deaths	Other Out-Flow	
Prior Month	Labor Force State	E	EE	EU	EN	**ED**	*EO*	EP
		U	UE	UU	UN	**UD**	*UO*	UP
		N	NE	NU	NN	**ND**	*NO*	NP
	In-Flow	Just 16	**JE**	**JU**	**JN**	0	*JO*	**JP**
		Other In-Flow	*IE*	*IU*	*IN*	0	0	*IP*
	Column Total		EC	UC	NC	**DC**	*OC*	Total

"Deaths" refers to those individuals who were in-scope in the prior month, but died by the current month. "Just 16" refers to those individuals who just turned 16 in the current month. The grayed portions of Table 2 correspond to estimates of Flows and Stocks that can be estimated using weighted CPS data. The remaining parts of the tables consist of indirectly estimated

residual totals that are used in order to construct a table which "adds up".

As noted, the four bolded row estimates JE, JU, JN, and JP can be directly computed from the CPS, using the known ages of respondents in the prior and current months, and their labor force status in the current month (note that JP is the stock population estimate of those who just turned 16 years of age as of the current month).

Deaths are reported in the CPS, but for various reasons are undercounted by nearly half, so to estimate flows out-of-scope due to death we need to take a less direct approach. To get a more accurate estimate, the total number of deaths each month are estimated using average death rates for each gender and race category derived from mortality tables published annually in "Vital Statistics of The United States" by the CDC's National Center for Health Statistics. These death rates are applied to the CPS data to estimate total deaths (DC in the table). Then the deaths are allocated among labor force states based on average allocation from historical CPS data, generating the bolded column estimates ED, UD, and ND.

There are three cells in Table 2 which are defined to be zero: those who would be classified as in-flows but were immediate out-flows due to death, and those who would be defined as simultaneously "other in-flows" and "other out-flows".

Following is a detailed discussion of the computation of final raked Gross Flows tables. In each step of the computation we refer back to the basic table structure as defined in Table 2. Tables are computed for Men and Women separately, and the process below is applied independently to Men and Women.

Margin adjustment step: Construct the Stock labor force estimates EP, UP, NP, and JP for the prior month using the sampling weight for the prior month, and the Stock labor force estimates EC, UC, NC for the current month using the sampling weight for the current month. Construct the death estimates ED, UD, and ND by first estimating the total number of deaths (the sum of ED, UD, and ND) by taking the prior month's total population and multiplying that by a death rate (obtained from Vital Statistics records) appropriate to the particular demographic group (again, either Men or Women). We then distribute the total death estimate among the three labor force estimates (ED, UD, and ND) based on annual average estimates of the proportion of deaths by labor force classification obtained from the CPS.

The adjustments so far yield a total of P0 = (EP + UP + NP + JP) - (ED + UD + ND) as a potential population in the scope of CPS in the current month. The actual in-scope population is P1 = EC + UC + NC. Usually, due to immigration, P1 will be greater than P0. In this case, we set the "Other In-Flow" total IP to equal the discrepancy (P1 - P0). Flow estimates (IE, IU, and IN) that sum to IP are created by allocating the discrepancy to labor force states according to their proportion of the current month population. The "Other Out-flow" column is set *to* zero. Some of the time, particularly when CPS population controls are adjusted downward, P0 will be greater than P1. In this case we set the "Other Out-Flow" total OP to equal (P0 - P1), allocate flow estimates (EO, UO, NO, and JO) by their proportion of the previous month population net of deaths (including JP), and set "Other In-flows" to zero.

Matching Step: Construct weighted counts of the 12 flow cells (EE, EU, EN, UE, UU, UN, NE, NU, NN, JE, JU, JN) using the sampling weight for the current month for those individuals who had a labor force status in both the prior and current month. These population totals are constructed for the particular demographic table of interest (either Men or Women). The weighted flow counts will be too small by approximately a factor of 25% due to the fact that approximately 75% of the sample overlaps form month-to-month. The initial iteration step below will correct for this undercount.

Iteration Step: The table constructed above is not entirely consistent, since not all the cells are guaranteed to add up to the appropriate row and column totals. In order to obtain consistency, iterative raking is performed. The death estimates column (ED, UD, ND, and DC), the Other Out-Flow column, and the Other In-Flow row are held fixed. The remaining interior cells of the table are raked by iterative proportional fitting to ensure additivity to the Stock estimates in the row and column totals.

Final Factor Calculation: The following is done for the production of Gross Flows weights. We plan to put these weights on our public-use data file to aid outside analysts in calculating gross flows. Compute factors for each of the estimates (EE, EU, EN, UE, UU, UN, NE, NU, NN, JE, JU, and JN) for each of the demographic tables by dividing the final "GF" estimate obtained in the Iteration Step by the weighted sample count "WGT" obtained prior to raking (using the current month sampling weights of matched persons). For example, if the ith male is in the EU cell:

$$GFW_{mi} = CPSFW_{mi} * GF_{mEU}/WGT_{mEU}$$

Conclusion

The proposed gross flows estimation procedure accounts for persons moving in and out of the labor force, applies a consistent weight for both prior and current month, and results in gross flows estimates that are compatible with monthly labor force stock estimates for both the prior and current months. Gross flows weights in the micro data file allow for summation of subgroup data consistent with overall totals.

We implemented the proposed procedure for data from 1990 through 2003. Table 3 shows a comparison of row percentages--flows as a percentage of the population in a labor force state the previous month--for the current procedure and our proposed procedure.[1] As can be seen, the average differences are fairly small, 5 percent or less in magnitude.

Table 3
Gross Flow Row Percentages, CPS 1990-2003
rel % diff = (Proposed-Current)/Current as a percent

	EE	EU	EN	UE	UU	UN	NE	NU	NN
Avg Row % Current	95.8	1.4	2.8	27.6	49.6	22.8	4.7	2.5	92.8
Avg Row % New	95.9	1.4	2.7	27.5	50.7	21.8	4.9	2.7	92.5
Rel % Diff	0	3	-3	0	2	-4	3	5	0

Our discussion above noted that rotation group bias caused the current procedure to show more substantial flows out of the labor force than could be reconciled with the stock data. Table 3 shows that the proposed procedure reduces flows out of the labor force as reflected in the relative difference figures for EN and UN.

A seasonally adjusted series of gross flows has also been developed. The non-seasonally adjusted series created using the proposed methodology is run through a seasonal adjustment procedure. The seasonally adjusted flows are then raked to seasonally-adjusted stocks using a raking procedure similar to the one described here. For more information on the seasonal adjustment of gross flows see "Analysis of Raking on Seasonally Adjusted Household Gross Flows Data" by Thomas Evans in these proceedings.

---

[1] Flows from January and April 1994 are excluded due to linking problems associated with a redesign of the survey.

In testing, several finer demographic breakouts were tried than the two-table (men and women) breakout in the proposal. Acceptable results were obtained for tables that were not seasonally adjusted. Difficulties encountered in seasonal adjustment spurred us to pare down the proposed demographic detail.

The proposed methods in this paper directly address only the additivity problems in gross flows estimation that arise due to non-matched cases, rotation group bias, and inconsistent weights. Other problems remain and are the topic of continuing research. For example, misclassifications tend to inflate the off-diagonal flows (EU, EN, UE, UN, NE, and NU). See "A Look at Several Gross Flows Estimation Methods with the Current Population Survey" by Steven Miller in these proceedings.

Disclaimer
Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics or the Bureau of the Census.

References.

Abowd, John M. and Zellner, Arnold. (1985). "Estimating Gross Labor-Force Flows," Journal of Business and Economic Statistics 3, 254-83.

Bureau of Labor Statistics and U.S. Census Bureau (2000). "Current Population Survey: Design and Methodology, Technical Paper 63." (www.census.gov/prod/2000pubs/tp63rv.pdf).

Evans, Thomas D. (2004). "Analysis of Raking on Seasonally Adjusted Household Gross Flows Data". Proceedings of the 2004 Joint Statistical Meetings.

Gordon, Robert A. (Chairman). (1962, September 27). Measuring Employment and Unemployment. Final Report of President's Committee to Appraise Employment and Unemployment Statistics.

Levitan, Sar A. (Chairman). (Labor Day, 1979). Counting the Labor Force. Final Report of National Commission on Employment and Unemployment Statistics.

Miller, Stephen M. (2004). "A Look at Several Gross Flows Estimation Methods with the Current Population Survey." Proceedings of the 2004 Joint Statistical Meetings.