

Quality of Match for Statistical Matches for the US

American Time Use Survey 2019, the Survey of Consumer Finances 2019, and the Annual Social and Economic Supplement 2020

This paper describes the quality of the statistical matching between the March 2020 supplement to the Current Population Survey (CPS) and the 2019 American Time Use Survey (ATUS) and Survey of Consumer Finances (SCF), which are used as the basis for the 2019 Levy Institute Measure of Economic Well-Being (LIMEW) estimates for the United States. In the first part of the paper, the alignment of the datasets is examined. In the second, various aspects of the match quality are described. The results indicate that the matches are of high quality, with some indication of bias in specific cases.

Data Alignment and Statistics

ATUS ?meta:atus.year – ASEC ?meta:asec.year

In order to create the synthetic dataset and transfer the time use information from the donor to the recipient dataset we need to verify that the distribution of information across surveys is comparable. To do so, we compare the distribution of individuals across the following set of variables: Gender, Parental Status, labor force Status, Presence of a Spouse, and employment status of the spouse. Since both datasets were collected within one year of each other, one should expect them to be well aligned.

Table 1: Summary Statistics, Alignment across Main variables

	ASEC	ATUS	diff
N	264,870,782	262,823,885	-0.8
Sex			
Female	51.4	51.6	0.2

	ASEC	ATUS	diff
Male	48.6	48.4	-0.2
Parental Status			
No	76.1	75.6	-0.5
Yes	23.9	24.4	0.5
Labor Force Status			
Not Employed	42.0	38.4	-3.6
Employed	58.0	61.6	3.6
Spouse present			
No	51.0	50.6	-0.4
Yes	49.0	49.4	0.4
Spouse's labor force status			
Spouse not Employed	18.1	17.3	-0.8
Spouse Employed	30.6	32.1	1.6

As can be observed in Table 1, the distribution of the sample with respect to sex and parental status is almost identical for both the ASEC and ATUS, with 48.5 percent of the sample being male, and about 24.1 percent being parents. The labor force status shows a relatively larger imbalance. The ATUS indicates there is a 3.6 percentage point larger share of employed individuals in the sample compared to the corresponding statistic in the ASEC survey (58 percent). The distribution of individuals across marital status presents a less severe imbalance. The statistics show that the share of married individuals is larger (3.6 percentage points) in the ATUS compared to the ASEC. In terms of the spouse's labor force status, the differences in the distribution among married individuals are negligible.

Table 2: Summary Statistics, Alignment across Selected variables

	ASEC	ATUS	diff
Household income category			
<\$15k	7.3	8.6	1.3
\$15k-\$35k	16.1	16.7	0.6
\$35k-\$50k	12.7	12.7	-0.0
\$50k-\$75k	18.6	19.3	0.7
>\$75k	45.3	42.7	-2.6
Age category			
15 to 24	15.8	16.2	0.4
25 to 34	17.1	16.9	-0.2
35 to 44	15.6	15.6	-0.1
45 to 54	14.9	15.3	0.3
55 to 64	15.9	16.1	0.2
65 to 74	12.2	11.8	-0.4

	ASEC	ATUS	diff
75 and older	8.4	8.1	-0.3
Race			
White	62.3	64.8	2.5
Black	12.0	11.9	-0.1
Hispanic	17.0	16.9	-0.2
Other	8.7	6.4	-2.2
Educational attainment			
Less than high school	13.9	14.4	0.4
High school	26.6	27.5	0.9
Some college	16.7	15.5	-1.2
College/grad school	42.8	42.6	-0.1
Number of children under 18 in household			
0	63.6	64.3	0.7
1	15.6	15.5	-0.1
2	13.2	12.4	-0.8
3	5.2	5.3	0.2
4 or more	2.5	2.4	-0.1
Number of persons in household over 18			
1	16.6	19.2	2.6
2	52.2	55.5	3.4
3	18.6	16.4	-2.3
4	8.5	7.0	-1.6
5 or more	4.0	2.0	-2.1

Table 2 presents statistics on additional variables that characterize the observations in both the donor and recipient datasets. The distribution across household income categories shows some imbalance, with the ATUS showing a considerably lower proportion of households in the highest income category, suggesting some undersampling of high-income households. For other demographic characteristics, such as age, race, and educational attainment, the distribution of individuals in both surveys is close. The largest observed differences across these characteristics are seen in the categories of some college (-1.2 percentage points) and whites (62.3 percentage points), with other differences falling below 1 percentage points. Finally, in terms of household structure, the survey’s distribution is close in terms of number of children in the household, with slightly larger discrepancies in terms of the number of adult persons in the household, where the ATUS indicates a smaller share of larger households.

As expected, although some differences in the distributions can be observed between both surveys, these differences are small and there are no systematic differences that might seriously affect the quality of the matching process.

SCF ?meta:scf.year – ASEC ?meta:asec.year

Similar to the previous case, in order to create the synthetic dataset that combines the SCF and ASEC information, we use five strata variables to evaluate the alignment across surveys. These strata variables are income category, homeownership, family type, and race and age of the householder. In this case, the households/families rather than individuals are used as the unit of observation. Table 3 presents summary statistics that compare the distribution of observations within the strata variables. Since both datasets were collected within one year of each other, one should expect them to be well aligned.

Table 3: Summary Statistics, Alignment across Main variables

	ASEC	SCF	diff
N	128,451,203	128,642,431	0.1
Household income category			
<\$20K	15.4	14.9	-0.5
\$20k-\$50k	25.7	30.7	5.0
\$50k-\$75k	15.6	17.1	1.6
\$75K-\$100k	11.6	11.3	-0.3
>\$100k	31.7	25.9	-5.8
Homeownership			
Renter	33.5	35.1	1.6
Owner w/Mortgage	36.6	42.1	5.5
Owner wo/Mortgage	29.9	22.8	-7.0
Family type			
Couple	55.1	56.0	0.9
Single Female	26.8	26.1	-0.7
Single Male	18.1	17.8	-0.2
Race category			
White	66.1	68.0	1.9
Black	12.8	15.7	2.9
Hispanic	13.7	10.9	-2.8
Other	7.4	5.4	-2.0
Age Category			
<35 and younger	19.2	20.9	1.6
35 to 49	24.8	24.7	-0.2
50 to 65	28.0	27.9	-0.1
>65 and older	27.9	26.6	-1.4

As observed in Table 3, the distribution of households across income categories shows good balance across both samples, displaying at most a 5.8 percentage point difference when comparing

households with highest income group. The SCF has a smaller share of middle-to-high-income households. This difference, however could be attributed to standard income growth across years. Nevertheless, for the matching, we aim to use income ranking rather than income levels to ensure good matching. Based on race and age, the distribution is very well balanced, with a less than 3 percentage point difference in the distributions, and a small underrepresentation of Hispanic and other races in the SCF. The largest distributional differences are present across family type and homeownership. The SCF dataset shows a slightly larger share of households within the “couples” categories (0.9 percentage points), while households with single females are underrepresented (-0.7 percentage points). Regarding homeownership, both samples present similar shares of renters and homeowners. Within the homeowners category, however, the ASEC underrepresents households with mortgages in about 5.5 percent of the instances compared to the SCF. To reduce the impact of the imbalance on the matching, we make use to include an indicator for “mortgage” in the imputation process.

Table 4 presents statistics on additional variables that characterize the observations in both datasets. Information on education and occupation categories corresponds to that of the householder. The surveys are well balanced in terms of the educational attainment of the householder, the number of persons within the household, and the occupational categories. Although some imbalance is observed regarding the share of individuals with a Highschool degree.

Table 4: Summary Statistics, Alignment across Selected variables

	ASEC	SCF	diff
Education category			
Less than high school	9.0	10.7	1.7
High school grad	27.2	24.5	-2.7
Some college	26.9	28.5	1.6
College or higher	36.9	36.3	-0.6
Sex of householder			
Male	72.7	73.0	0.3
Female	27.3	27.0	-0.3
Number of members in household			
1 Member	28.2	27.2	-1.0
2 Members	34.8	35.1	0.3
3 Members	15.1	16.1	1.0
4 or more Member	21.9	21.6	-0.3
Occupation category			
Occ1: 37–199	29.3	29.5	0.2
Occ2: 203–389	10.7	11.6	0.9
Occ3: 403–469 & 903–905	7.4	8.9	1.5
Occ4: 503–699	9.9	10.0	0.1

	ASEC	SCF	diff
Occ5: 703–889	7.5	7.4	-0.1
Occ6: 473–499	0.7	1.2	0.6
Other	34.5	31.3	-3.3