Rios-Avila, Fernando

Working Paper

Quality of match for statistical matches using the American Time Use Survey 2010, the Survey of Consumer Finances 2010, and the annual social and economic supplement 2011

Working Paper, No. 798

Provided in Cooperation with:
Levy Economics Institute of Bard College

Suggested Citation: Rios-Avila, Fernando (2014) : Quality of match for statistical matches using the American Time Use Survey 2010, the Survey of Consumer Finances 2010, and the annual social and economic supplement 2011, Working Paper, No. 798, Levy Economics Institute of Bard College, Annandale-on-Hudson, NY

This Version is available at:
https://hdl.handle.net/10419/110010

Working Paper No. 798

Quality of Match for Statistical Matches Using the American Time Use Survey 2010, the Survey of Consumer Finances 2010, and the Annual Social and Economic Supplement 2011

by

Fernando Rios-Avila*
Levy Economics Institute of Bard College

May 2014

* The author thanks Thomas Masterson for his guidance in completing the matching process and Ajit Zacharias for the helpful comments on this paper.

Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
http://www.levyinstitute.org

ISSN 1547-366X

ABSTRACT

This paper describes the quality of the statistical matching between the March 2011 supplement to the Current Population Survey and the 2010 American Time Use Survey and Survey of Consumer Finances, which are used as the basis for the 2010 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.

Keywords: Statistical Matching; American Time Use Survey; Survey of Consumer Finances; LIMEW; United States

JEL Classifications: C14, C40, D31

INTRODUCTION

This paper describes the construction of the synthetic dataset created for use in the estimation of the Levy Institute Measure of Economic Well-Being (LIMEW) for the United States. LIMEW was developed as an alternative to conventional income measures that provides a more comprehensive measure of economic well-being.[1] Construction of the LIMEW requires a variety of information for households. In addition to the standard demographic and household income information, the estimation process also requires information about household members time use and information on household's wealth, assets and debts. Unfortunately, no single dataset contains all required data for the estimation.

In order to produce LIMEW estimates, a synthetic dataset is created combining information from three datasets, applying a statistical matching process.[2] For the United States, the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) 2011 is used as the base dataset, as it contains good information regarding demographic, social and economic characteristics, as well as income, work experience, noncash benefits and migration status of persons 15 years old and over. Time use data comes from the American Time Use Survey (ATUS) 2010, which provides rich data regarding how people divide their time among life's activities, including time spent doing paid and unpaid activities, inside and outside the household. Wealth data come from the Survey of Consumers Finances (SCF) 2010, which collects detailed information on household finances, income, assets and liabilities.
This paper is organized as follows. Section one describes the data. Section two assesses the alignment of the information between ASEC and ATUS, at the individual level, and the ASEC and the SCF at the household level. Section three briefly describes the methodology and analyzes the matching quality of the statistical matching. Section four concludes.

[1] For details on the background of the LIMEW see Wolff and Zacharias (2003).
[2] For further details on the methodology see Kum and Masterson (2010).

1. DATA DESCRIPTION

1.1. Annual Social Economics Supplement (ASEC)

The Current Population Survey (CPS) is a monthly survey administered by the U.S. Bureau of Labor Statistics. It is used to assess the activities of the population and provide statistics such as employment, and unemployment on the current labor market. Each household in the CPS is interviewed for four consecutive months, not interviewed for eight, and interviewed again for four additional months. Although the main purpose of the survey is to collect information on the labor market situation, the survey also collects detailed information of demographic characteristics (age, sex, race, and marital status), educational attainment and family structure.

In March of every year, the households previously interviewed answer additional questions, part of the Annual Social Economic (ASEC) supplement, formerly known as the Annual Demographic File. In addition to the basic monthly information, this supplement provides additional data on work experience, income, noncash benefits and migration.

The ASEC 2011 is used as the base dataset (recipient), as it contains rich information regarding demographics and economic status. Because the time-use survey (described below) covers individuals of 15 years of age and older, younger individuals are discarded from the ASEC sample. This leaves us with a total of 156,748 observations, representing 243,803,280 individuals when weighted. For the household level analysis, only information regarding the householder is used, leaving 75,148 observations, representing 118,682,616 households when weighted.

1.2. American Time Use Survey (ATUS)

The American Time Use Survey (ATUS), a survey sponsored by the Bureau of Labor Statistics and Collected by the U.S. Census Bureau, is the first continuous survey on time use in the United States available since 2003. Its main objective is to provide nationally representative estimates of people's allocation of time among different activities, collecting information on what they did, where they were, and with whom they were.

The ATUS is administered to a random sample of individuals selected from a set of eligible households that have completed their final month interviews for the Current Population Survey (CPS). The ATUS covers all residents that are at least 15 years old, and are part of the civilian and non-institutional population in the United States.

The ATUS 2010, which contains a total of 13,260 observations, is used as the donor dataset to obtain information regarding time use, which will be transferred to the ASEC 2011. Since information regarding household income is incomplete, the information was imputed using a univariate imputation process and information from the ASEC-CPS 2010. After the imputation procedure, three records were left unmatched and eliminated from the sample, leaving a total of 13,257 observations, which represent 241,823,036 individuals when weighted.

1.3. Survey of Consumer Finances (SCF)

The Survey of Consumer Finances (SCF) is normally a triennial cross-sectional survey, sponsored by the Board of Governors of the Federal Reserve System in cooperation with the U.S. Department of Treasury, which collects information on families' balance sheets, pensions, income, and demographic characteristics.[3] The purpose of the survey is to provide detailed information on households' assets and liabilities that can be used for analyzing households' wealth and their use of financial services.

In order to provide reliable information on household wealth distribution, the SCF is based on a dual-frame sample design. On the one hand, a geographically based random sample is interviewed to obtain consistent information on attributes broadly distributed across the population. On the other hand, a supplemental sample was obtained to include a sample of wealthy families, to provide accurate information on wealth distribution, as the value of non-home assets and liabilities are highly concentrated. In order to deal with the missing data, most variables with missing values are imputed using a multiple imputation procedure from which five replicates (imputations) for each record are obtained. [4]

The SCF 2010 is used as the donor dataset to obtain information regarding assets, debts and net worth wealth. For the SCF 2010, a total of 6,492 families/households were interviewed. In order to account for the multiple imputation information, the five replicates are combined and used for the matching procedure. This provides a sample of 32,410 observations, representing 117,609,227 households when weighted.

[3] Over the 1983–1989 and 2007–2009 periods, the SCF has collected information in panel data.
[4] For information regarding the use and estimation replicate samples see Kennickell (2000) and Kennickell, Arthur B. and R. Louise Woodburn (1999).

2. DATA ALIGNMENT AND STATISTICS

2.1. ATUS 2010 – ASEC 2011

In order to create the synthetic dataset and transfer the time use information from the donor to the recipient dataset as closely as possible, five strata variables are used to perform the match within the defined subsamples (cells). These strata variables are *sex, parental status, labor force status, marital status* and *spouse's labor force status.* The combination of these five strata variables provides a total of 24 cells which are used to perform a within-cell match. Table 1 presents summary statistics that compare the distribution of individual within the strata variables. Since both datasets were carried out within one year of each other, one should expect them to be well aligned.

Table 1 Summary Statistics. Alignment Across Strata Variables

	ASEC	**ATUS**	**diff (%)**
Individuals	243,803,280	241,823,036	-0.8%
Sex			
Female	51.3%	51.5%	0.2%
Male	48.7%	48.5%	-0.2%
Parental Status			
No	63.2%	63.5%	0.3%
Yes	36.8%	36.5%	-0.3%
Labor Force Status			
Not employed	43.2%	38.9%	-4.3%
Employed	56.8%	61.1%	4.3%
Spouse			
No	45.17%	43.70%	-1.47%
Yes	54.83%	56.30%	1.47%
Spouse's Labor Force Status			
Spouse not employed	36.15%	36.00%	-0.15%
Spouse employed	63.85%	64.00%	0.15%

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

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 2011 and ATUS 2010, with 48.5% of the sample being male, and about 36.5% being parents. The labor force status shows a relatively larger imbalance. The ATUS indicates there is a 4.3 percentage point (pp) larger share of individuals employed in the sample compared to the corresponding statistic in the ASEC survey (56.8%). The distribution of individuals across marital status presents a less severe imbalance. The statistics show that the share of married individuals is larger (1.5pp) in the ATUS compared to the ASEC. In terms of the spouse Labor Force Status, the differences in the distribution of among married individuals are negligible.

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 with the highest income category, suggesting some under-sampling 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 are seen in the categories of Some College (1.8pp), and Whites (1.3pp), although both fall below 2pp. Finally, in terms of household structure, the surveys distribution is close in terms of number of children in the household, with slightly less favorable results in terms of 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. Based on the strata variables described above, a new variable is created with 24 matching cells.

Table 2 Summary Statistics. Alignment Across Selected Variables

	ASEC	ATUS	diff (%)
HH Income Category			
0-14999	10.37%	11.75%	1.38%
15000-34999	20.20%	22.09%	1.89%
35000-49999	13.79%	14.10%	0.31%
50000-74999	18.77%	19.97%	1.20%
75000+	36.86%	32.08%	-4.78%
Age category			
15 to 24	17.39%	17.36%	-0.03%
25 to 34	17.04%	16.92%	-0.12%
35 to 44	16.34%	16.57%	0.23%
45 to 54	18.02%	18.31%	0.29%
55 to 64	15.16%	14.84%	-0.32%
65 and older	16.06%	16.00%	-0.06%
Race			
White	67.28%	68.57%	1.29%
Black	11.68%	11.61%	-0.07%
Other	6.67%	5.55%	-1.12%
Hispanic	14.38%	14.27%	-0.11%
Educational Attainment			
Less than HS	17.75%	18.17%	0.42%
HS	28.89%	29.17%	0.28%
Some College	18.58%	16.83%	-1.75%
College/Grad school	34.77%	35.83%	1.06%
Number of children under 18 in household			
0	59.57%	58.93%	-0.64%
1	17.54%	17.37%	-0.17%
2	14.08%	14.46%	0.38%
3	5.86%	6.30%	0.44%
4	1.99%	1.97%	-0.02%
5 or more	0.95%	0.96%	-0.01%
Number of persons in household over 18			
0	0.01%	0.02%	0.01%
1	16.97%	18.85%	1.88%
2	52.79%	54.90%	2.11%
3	18.28%	16.85%	-1.43%
4	8.17%	7.14%	-1.03%
5	2.57%	1.36%	-1.21%
6 or more	1.21%	0.88%	-0.33%

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

2.2. SCF 2010 – ASEC 2011

Similar to the previous case, in order to create the synthetic dataset that combines the SCF and ASEC information, five strata variables are used to perform the statistical matching. These strata variables are Income Category, Home Ownership, Family Type, and Race and Age of the Householder (head of household). In this case, the households/families, rather than individuals, are used as a unit of observation. The combination of these five strata variables provides a total of 360 cells which are initially used to perform the match. Table 3 presents summary statistics that compare the distribution of observations within the strata variables. Since both datasets were carried out within one year of each other, one should expect them to be well aligned.

Table 3 Summary Statistics. Alignment Across Strata Variables

	ASEC	**SCF**	**diff (%)**
Individuals	118,682,616	117,609,227	-0.90%
HH Income Category			
lt $20k	19.8%	20.9%	1.1%
$20-50k	30.6%	33.7%	3.1%
$50-75k	17.7%	17.3%	-0.5%
$75-100k	11.4%	9.9%	-1.5%
gt $100k	20.4%	18.2%	-2.2%
Home Ownership			
Renter	33.8%	32.8%	-1.0%
Owner w/mortgage	39.8%	47.0%	7.2%
Owner wo/mortgage	26.4%	20.2%	-6.2%
Family Type			
Couple	54.6%	58.1%	3.5%
Single Female	27.9%	26.5%	-1.4%
Single Male	17.6%	15.4%	-2.1%
Race Category			
White	70.3%	70.8%	0.5%
Black	12.3%	13.8%	1.5%
Other	5.9%	4.6%	-1.2%
Hispanic	11.5%	10.8%	-0.7%
Age Category			
<35	21.7%	21.0%	-0.7%
35-49	28.2%	28.4%	0.2%
50-65	28.8%	28.4%	-0.3%
>=65	21.4%	22.2%	0.9%

Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

As observed in Table 3, the distribution of households across income category shows good balance across both samples, displaying at most a 3.1pp difference. The SCF has slightly smaller share of middle-to-high-income households. Based on race and age, the distribution is very well balanced, with less than 1.5pp difference in the distributions, with a small underrepresentation of Hispanic and other races in the SCF.[5] The largest distributional differences are present across family type and home ownership. The SCF dataset shows a larger share of households within "couples" categories (3.5pp), while households with single males are underrepresented (2.1pp).[6] Regarding homeownership, both samples present similar shares of renters and homeowners. Within homeowners, however, the ASEC underrepresents households with mortgages in about 7%, compared to the SCF. Under the assumption the ASEC information is correct, the excess of mortgage debt is redistributed among householders with mortgages. This strategy has the advantage of maintaining the total amount of mortgage debt unchanged in the imputed data, although this might imply some overestimation of the mortgage debt when comparing households with mortgage in both datasets (see Figure 6).

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 education attainment of the householder, and the number of persons within the household, and occupational categories.

[5] While the table shows the distribution for 4 age categories, the strata variable only differentiates between household heads older and younger than 65.

[6] It is possible that the underrepresentation of "couple" households are underrepresented in the ASEC survey, compared to the SCF, because the latter uses the definition of consumer unit, which is compared with the former "household" definition. In the ASEC definition, a household can contain more than one family (couple).

Table 4 Summary Statistics. Alignment Across selected Variables

	ASEC	SCF	diff (%)
Education Category			
Less than HS	12.4%	12.0%	-0.4%
high school grad	29.8%	32.2%	2.4%
College	27.4%	24.9%	-2.5%
More than College	30.5%	31.0%	0.5%
Sex of HH head			
Female	30.6%	27.1%	-3.4%
Male	69.5%	72.9%	3.4%
No. Persons in Household			
1 person	27.6%	25.6%	-2.0%
2 persons	33.5%	32.9%	-0.6%
3 or more	39.0%	41.5%	2.6%
Occupation Category			
Occ1: 37-199	26.6%	27.7%	1.2%
Occ2: 203-389	13.7%	11.9%	-1.8%
Occ3: 403-469&903-905	8.8%	9.9%	1.1%
Occ4: 503-699	10.8%	10.3%	-0.5%
Occ5: 703-889	7.4%	7.7%	0.3%
Occ6: 473-499	0.6%	0.9%	0.3%
Other	32%	31.7%	-0.6%

Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

The distribution of *Sex* of the household head shows some imbalance across both datasets. In the ASEC, the householder or person of reference is selected randomly in cases of couples. For consistency, we assign the male within the couple to be considered as the head of the household, a definition closer to the SCF head of the household. While the SCF survey indicates that a large share of households (72.9%) are male, the ASEC shows 69% of householders to be male. Although this difference is relatively large compared to the one observed across other characteristics (about 3%), it should not have an effect on the quality of the matching. The next section describes the quality of the matching.

3. Matching Quality

Statistical matching is a widely used technique, predominantly in observational studies in the medical literature. This method consists of combining the information of two separate and independent surveys into a single combined dataset from which statistical inferences can be obtained. The methodology enables the combination of the datasets using common information between both surveys, preserving the distributional characteristics of the combined information.[7] In the following, the match quality between the ASEC 2011 and ATUS 2010 and SCF 2010, correspondingly, are assessed.

3.1. Matching: ATUS and ASEC

In order to obtain a good match, the matching process begins using five strata variables, namely Sex, Parental Status, Labor Force Status, Marital Status and Spouse Labor Force Status, to obtain 24 matching cells. Within each of these cells, propensity scores are estimated using logit models. A dummy variable indicating if the observation corresponds to the donor or the recipient survey is used as a dependent variable. A set of demographic variables (i.e. age, educational attainment, race, parental status, marital status and employment status) and household characteristics (i.e. number of adults, number of children, household monthly income) are included as independent variables. For subsequent matching rounds, broader matching cells are defined accordingly, keeping the logit specifications consistent across all models, including the omitted strata variable in the specification. The logit models and propensity scores are estimated using all information within broader cells, but the matching is elaborated only across observations left unmatched from previous rounds.

Turning to the results of the match performance, Table 5 shows the distribution of the matched records by matching round. As expected from these types of processes, a large share of the matches (95.2%) occurs on the first round, which ensures the highest level of quality match. At the same time, only 0.2% of the weighted sample was left unmatched after seven matching rounds. These unmatched observations should not bias the distributional statistics of the transferred information.

[7] For further details on the matching procedure see Kum and Masterson (2010).

Table 5 Distribution of Matched Records by Matching Round

Matching Round	Records Matched	Percent	Cumulative Percent
1	231,982,730	95.2	95.2
2	1,326,694	0.5	95.7
3	6,382,782	2.6	98.3
4	1,510,943	0.6	98.9
5	323,024	0.1	99.1
6	1,347,063	0.6	99.6
7	346,949	0.1	99.8
8	583,095	0.2	100
Total	243,803,280	100	

Source: Author's calculations based on ASEC 2011 and ATUS 2010 matched data

Table 6 provides a description of the match quality, comparing some distributional statistics on the weekly hours of household production between the original information (ATUS) and the imputed data (ASEC). Table 6 also presents some statistics on three components of household production.[8] Given the large presence of zero hours allocated to household production in the sample, some ratios and statistics are not available. The percentile ratios are all equivalent with identical Gini coefficients (0.5223). The means and medians on the disaggregated components of Household production also show a strong equivalence between both surveys, indicating a strong balance in aggregate terms.

[8] Household production can be broadly categorized in three groups or components: care (child care, education, etc), procurement (shopping, etc) and core (cooking, cleaning, laundry, etc).

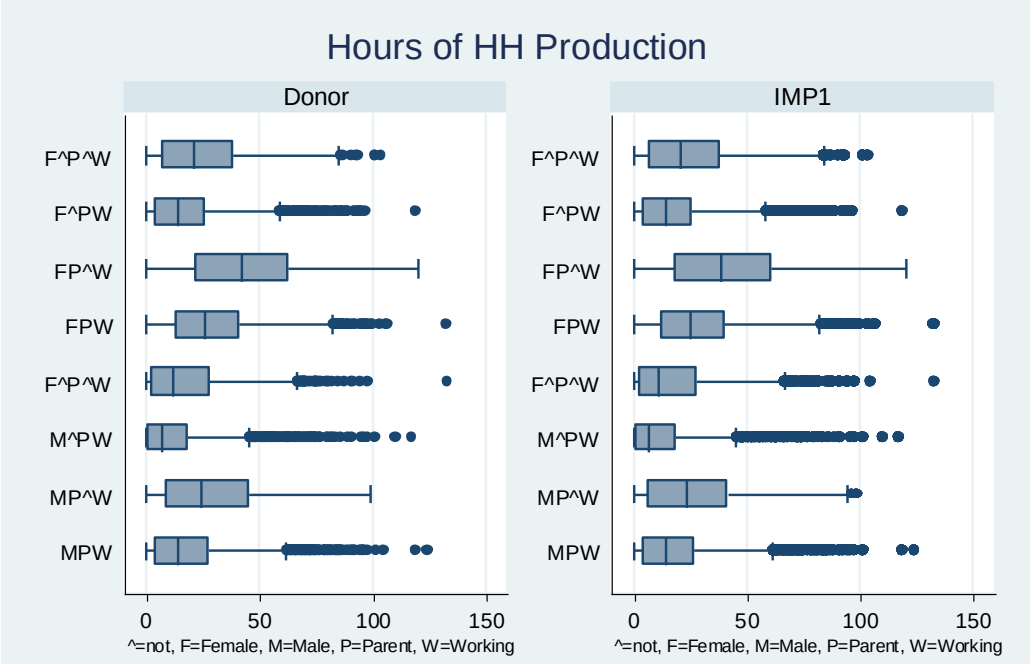
Table 6 Matching Quality: Summary Statistics

	ATUS	ASEC	Ratio ASEC/ATUS
Distributional Statistics (HH Production Wkly Hrs)			
p90/p10	n/a	n/a	
p90/p50	3.36	3.36	100%
p50/p10	n/a	n/a	
p75/p25	7.94	7.97	100%
p75/p50	2.09	2.09	100%
p50/p25	3.81	3.81	100%
Gini	0.52	0.52	100%
Summary Statistics			
Average HH Production Wkly Hrs	22	22	100%
Average Care Wkly Hrs	3.6	3.6	100%
Average Procurement Wkly Hrs	5.2	5.2	100%
Average Core Wkly Hrs	13	13	100%
Median HH Production Wkly Hrs	16	16	100%
Median Care Wkly Hrs	n/a	n/a	
Median Procurement Wkly Hrs	n/a	n/a	
Median Core Wkly Hrs	7	7	100%

Note: Household production activities are classified in three classes: care, such as child care and education; procurement, such as shopping groceries and clothes; and core, such cooking and cleaning.
Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

Figure 1 presents a visual representation of the distribution of hours allocated to Household production using three of the strata variables: Sex, Parental Status, and Labor Force Status. The figure shows that except for some values on the right tail of the distributions, for example women who are not parent and are not working (F^P^W) Men, who are parents and are not working (MP^W), the overall distributions within the strata variables are analogous indicating a good quality of the match.

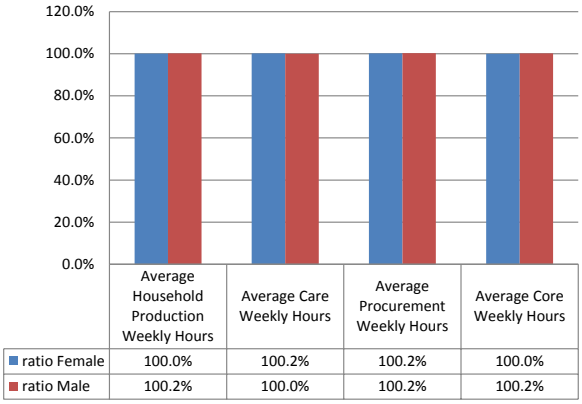
Figure 1 Distritubtion of Hours in Household Production, by Survey



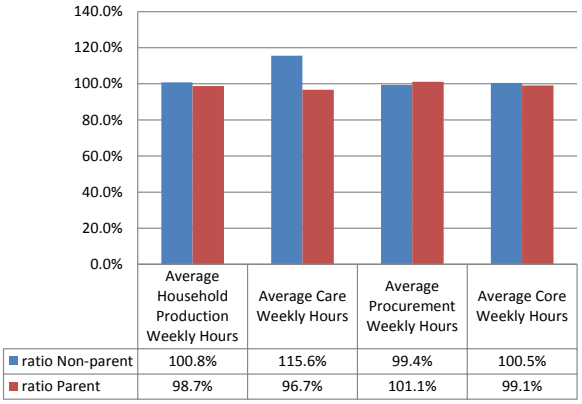
Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

For a detailed review of the performance of the matching, Figure 2 shows the ratios of the disaggregated hours allocated to household production (care, procurement and core) between the imputed data (ASEC) and the donor data (ATUS). Table 6 provides additional information on the mean and median of hours of household production per week. The information is shown across the five strata variables used for the matching. With some exceptions, the ratios of mean weekly hours of household production (and subcategories) fall within 2% of difference across all strata variables, an indication of good match quality. The largest differences are observed among non-parents and unmarried people. In both cases the statistics indicate 15.6% and 7.0% more hours allocated to care activities in imputed data, and, as it can be observed on Table 7, total hours of household production for these particular groups differ only in 0.2 hours for both means and medians. In perspective, while such differences seem large, they might have a small effect on other analyses since the average hours allocated to care among the specific groups are rather small (1.7 and 0.8 hours). Finally, when looking at the labor status variables, while all ratios fall within tolerable limits, the imputed ASEC information underestimates all household production hours among nonworking individuals, which translates into a little less than a 1-hour difference between the imputed and donor information (see Table 7).

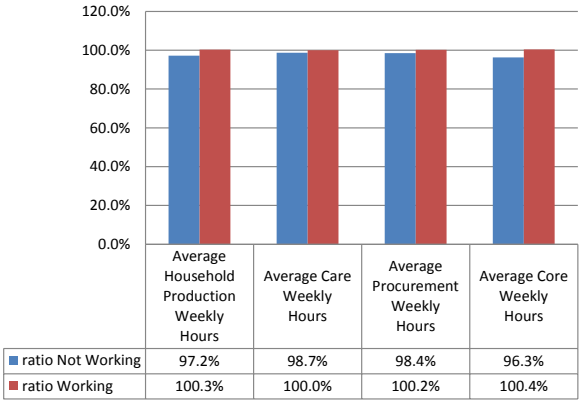
Figure 2 Ratio of Mean Household Production Hours, by Strata Variables



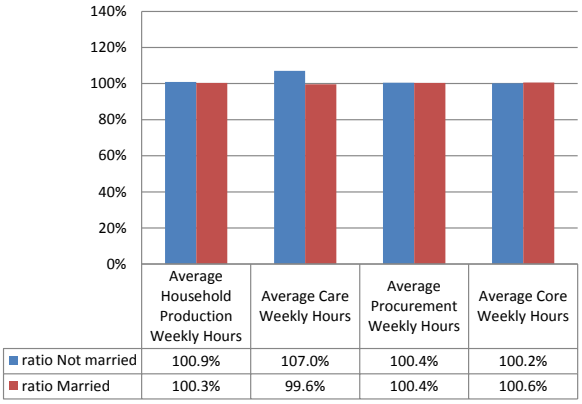
	Average Household Production Weekly Hours	Average Care Weekly Hours	Average Procurement Weekly Hours	Average Core Weekly Hours
ratio Female	100.0%	100.2%	100.2%	100.0%
ratio Male	100.2%	100.0%	100.2%	100.2%



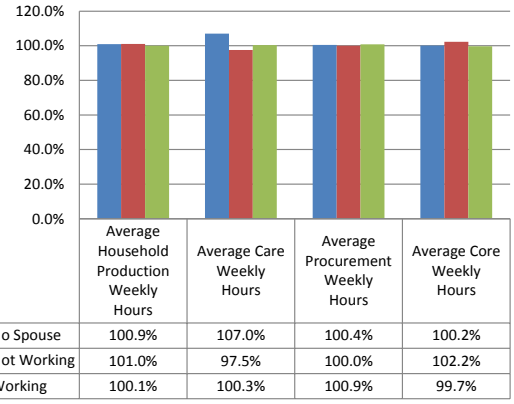
	Average Household Production Weekly Hours	Average Care Weekly Hours	Average Procurement Weekly Hours	Average Core Weekly Hours
ratio Non-parent	100.8%	115.6%	99.4%	100.5%
ratio Parent	98.7%	96.7%	101.1%	99.1%



	Average Household Production Weekly Hours	Average Care Weekly Hours	Average Procurement Weekly Hours	Average Core Weekly Hours
ratio Not Working	97.2%	98.7%	98.4%	96.3%
ratio Working	100.3%	100.0%	100.2%	100.4%



	Average Household Production Weekly Hours	Average Care Weekly Hours	Average Procurement Weekly Hours	Average Core Weekly Hours
ratio Not married	100.9%	107.0%	100.4%	100.2%
ratio Married	100.3%	99.6%	100.4%	100.6%



	Average Household Production Weekly Hours	Average Care Weekly Hours	Average Procurement Weekly Hours	Average Core Weekly Hours
ratio No Spouse	100.9%	107.0%	100.4%	100.2%
ratio Not Working	101.0%	97.5%	100.0%	102.2%
ratio Working	100.1%	100.3%	100.9%	99.7%

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

Table 7 Average and Median, Household Production Weekly Hours, by Selected Variables

	Averages			Median		
	Donor	**Imputed**	**Ratio**	**Donor**	**Imputed**	**Ratio**
HH production	22	22	100.0%	16.0	16.0	100.0%
Care	3.6	3.6	100.3%	0.0	0.0	
Procurement	5.2	5.3	100.2%	0.0	0.0	
Core	13.2	13.2	100.0%	7.0	7.0	100.0%
Marital Status						
Not married	17	17.2	100.9%	10.9	11.1	102.1%
Married	25.9	26	100.3%	20.5	20.9	101.7%
Parenthood						
Non-parent	18.3	18.5	100.8%	12.6	12.8	101.8%
Parent	28.5	28.1	98.7%	23.5	23.2	99.0%
Sex						
Female	26.5	26.5	100.0%	21.6	21.6	100.0%
Male	17.2	17.3	100.2%	10.7	10.9	101.1%
Labor Status						
Not Working	26.4	25.6	97.2%	21.0	20.4	97.2%
Working	19.3	19.3	100.3%	14.0	14.0	100.0%
Spouse Labor Status						
No Spouse	17	17.2	100.9%	10.9	11.1	102.1%
Not Working	23.6	23.8	101.0%	18.7	18.7	100.0%
Working	27.2	27.3	100.1%	21.7	21.7	100.0%
Education						
Less than HS	18.2	21	115.3%	11.1	14.6	131.6%
High school	23.1	22.6	97.5%	17.5	16.3	93.3%
Some College	21.7	21.7	99.9%	16.3	15.8	96.4%
College Grad	23.2	22.3	96.1%	17.5	16.3	93.3%
HH Income						
0-14,999	21.7	20.3	93.4%	15.2	14.0	92.3%
15,000-34,999	23.0	21.7	94.2%	17.5	15.2	86.7%
35,000-49,999	22.2	22.3	100.2%	15.8	16.3	103.7%
50,000-74,999	21.4	22.3	104.2%	15.8	16.3	103.7%
75,000+	21.8	22.5	103.4%	14.0	16.9	120.9%
Age Group						
15 to 24	14.3	17.2	120.6%	7.0	10.5	150.0%
25 to 34	22.5	22.6	100.5%	15.8	16.9	107.4%
35 to 44	27.0	24.7	91.4%	21.2	18.7	87.9%
45 to 54	22.6	23.1	102.2%	16.9	17.4	102.7%
55 to 64	22.1	21.3	96.4%	16.9	15.2	89.7%
65 and older	24.0	23.3	97.1%	20.4	18.1	88.5%

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

Table 8 presents additional details on the quality of the match using the cell matching variable. Similar to the results described before, with some exceptions, total household production—in particular procurement and core hours—shows good levels of balance across most of the matching cells (note: Procurement and Core hours are part of household production). Within cells 8 and 19, however, the imputed sample underestimates the allocation of hours in household production, particularly from core activities, in almost 15% (2-5 hours). These cells are the ones that had the lowest rate of first round matching, which could explain these results. Allocation of time to "care" shows more imbalances across the matching cells compared to the other imputed variables, with many cells showing more than 5% differences. Cells 8 and 19 present a considerable underestimation of about 3.2 hours and 1.1 hours on care activities in the imputed ASEC data, but they represent less than 2% of the sample. While cells 4, 13 and 15 also present large relative imbalances, the absolute differences are negligible (less than 0.5 hours).

Table 8 Ratio and Absolute Differences of Mean Household Production Hours, by Matching Cell

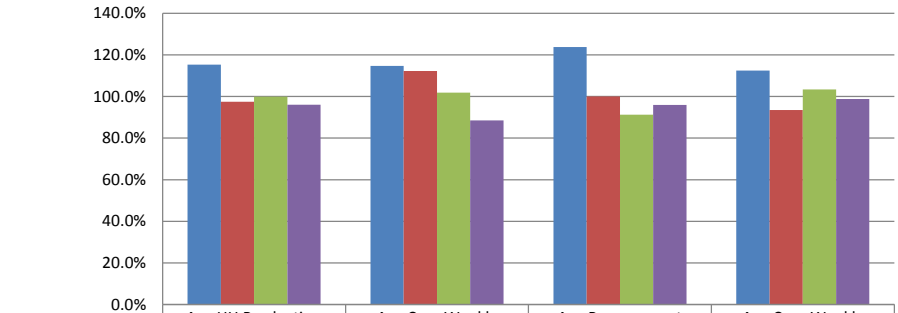
Cell	Sex	Parent Status	Labor Status	Spouse Status	Average Household Production Weekly Hours ratio(abs diff)	Average Care Weekly Hours ratio(abs diff)	Average Procurement Weekly Hours ratio(abs diff)	Average Core Weekly Hours ratio(abs diff)
C1	W	N	Not working	No	99%(0.2hrs)	108%(0.1hrs)	100%(0.0hrs)	98%(0.3hrs)
C2	W	N	Not working	Not working	101%(0.3hrs)	123%(0.2hrs)	99%(0.1hrs)	101%(0.1hrs)
C3	W	N	Not working	Working	100%(0.1hrs)	102%(0.0hrs)	100%(0.0hrs)	100%(0.1hrs)
C4	W	N	Working	No	102%(0.3hrs)	88%(0.1hrs)	102%(0.1hrs)	103%(0.3hrs)
C5	W	N	Working	Not working	100%(0.1hrs)	103%(0.0hrs)	92%(0.5hrs)	103%(0.5hrs)
C6	W	N	Working	Working	99%(0.1hrs)	97%(0.0hrs)	101%(0.1hrs)	99%(0.2hrs)
C7	W	Y	Not working	No	95%(1.6hrs)	92%(0.6hrs)	98%(0.1hrs)	95%(0.9hrs)
C8	W	Y	Not working	Not working	87%(5.5hrs)	74%(3.2hrs)	108%(0.4hrs)	88%(2.7hrs)
C9	W	Y	Not working	Working	98%(0.9hrs)	98%(0.4hrs)	99%(0.1hrs)	99%(0.4hrs)
C10	W	Y	Working	No	93%(1.9hrs)	90%(0.7hrs)	94%(0.4hrs)	94%(0.8hrs)
C11	W	Y	Working	Not working	101%(0.3hrs)	101%(0.0hrs)	96%(0.2hrs)	103%(0.4hrs)
C12	W	Y	Working	Working	100%(0.0hrs)	101%(0.1hrs)	103%(0.2hrs)	98%(0.3hrs)
C13	M	N	Not working	No	101%(0.1hrs)	170%(0.5hrs)	100%(0.0hrs)	96%(0.4hrs)
C14	M	N	Not working	Not working	100%(0.0hrs)	100%(0.0hrs)	97%(0.2hrs)	101%(0.2hrs)
C15	M	N	Not working	Working	98%(0.5hrs)	142%(0.3hrs)	99%(0.0hrs)	96%(0.7hrs)
C16	M	N	Working	No	102%(0.3hrs)	95%(0.0hrs)	102%(0.1hrs)	102%(0.2hrs)
C17	M	N	Working	Not working	95%(0.8hrs)	100%(0.0hrs)	92%(0.4hrs)	95%(0.4hrs)
C18	M	N	Working	Working	100%(0.0hrs)	102%(0.0hrs)	100%(0.0hrs)	100%(0.0hrs)
C19	M	Y	Not working	No	87%(3.2hrs)	76%(1.1hrs)	96%(0.2hrs)	87%(1.9hrs)
C20	M	Y	Not working	Not working	96%(1.0hrs)	90%(0.7hrs)	114%(0.8hrs)	92%(1.2hrs)
C21	M	Y	Not working	Working	98%(0.6hrs)	101%(0.1hrs)	98%(0.1hrs)	97%(0.5hrs)
C22	M	Y	Working	No	97%(0.5hrs)	92%(0.3hrs)	100%(0hrs)	99%(0.2hrs)
C23	M	Y	Working	Not working	100%(0.1hrs)	96%(0.3hrs)	98%(0.1hrs)	103%(0.3hrs)
C24	M	Y	Working	Working	99%(0.2hrs)	99%(0.0hrs)	101%(0.0hrs)	99%(0.1hrs)

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

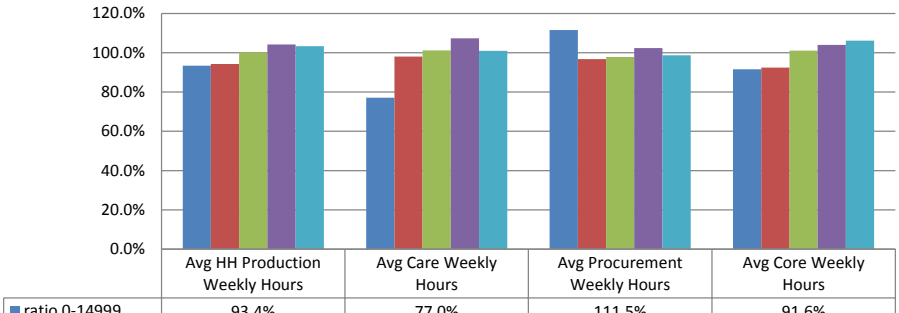
To examine the match quality beyond the framework of the strata variables, Figure 3 presents information on ratios for the household production and its components across education, household income level, and age group. In addition, Table 7 provides the mean and median of total household production for selected variables. In terms of education, people with high school and some college education have good levels of balance between both surveys. People with less than high school education are imputed with longer hours allocated to household production (2.8 hours more) and all its components. In contrast, there is a consistent underestimation of the hours of household production (0.9 hours) for people with at least a college degree. Individuals in the lowest two income groups show an underestimation of the hours allocated to HH Production (1.3-1.4 hours), a bias that is particularly large when observing at the hours assigned to care and core activities. Similar gaps are observed when looking at the medians.

In terms of age groups, while overall hours of household production seem to be adequately balanced across the samples, with the exception of the youngest group, the disaggregated components show large unbalanced ratios, especially for people 55 or older. In the imputed sample, total number of hours in household production corresponding to individuals 15-24 years of age is in average about 3 hours above than that in the ATUS estimates, with similar bias when considering the medians, which comes directly from the overestimation of core household activities. Information regarding care activities presents the largest imbalances across all age groups. People between 15 and 44 years old present an underestimation of 0.8 to 2.7 lower hours assigned to care activities. In contrast, for people 45 years old or older, hours assigned to care activities are overestimated by about 1.5 to 1.9 hours, representing up to a 280% over-estimation.

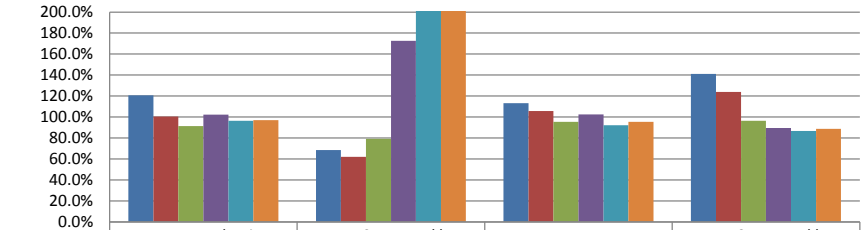
Figure 3 Ratio of Mean Household Production Hours, by Selected Variables



	Avg HH Production Weekly Hours	Avg Care Weekly Hours	Avg Procurement Weekly Hours	Avg Core Weekly Hours
ratio Less than HS	115.3%	114.7%	123.7%	112.4%
ratio Highschool	97.5%	112.2%	100.0%	93.5%
ratio Some College	99.9%	101.8%	91.3%	103.4%
ratio College Grad	96.1%	88.4%	95.9%	98.8%



	Avg HH Production Weekly Hours	Avg Care Weekly Hours	Avg Procurement Weekly Hours	Avg Core Weekly Hours
ratio 0-14999	93.4%	77.0%	111.5%	91.6%
ratio 15000-34999	94.2%	98.1%	96.7%	92.5%
ratio 35000-49999	100.2%	101.2%	97.8%	101.0%
ratio 50000-74999	104.2%	107.3%	102.4%	104.0%
ratio 75000+	103.4%	101.0%	98.7%	106.1%



	Avg HH Production Weekly Hours	Avg Care Weekly Hours	Avg Procurement Weekly Hours	Avg Core Weekly Hours
Ratio 15-24	120.6%	68.6%	113.1%	141.1%
Ratio 25-34	100.5%	62.0%	105.7%	123.8%
Ratio 35-44	91.4%	79.2%	95.3%	96.4%
Ratio 45-54	102.2%	172.6%	102.5%	89.4%
Ratio 55-64	96.4%	296.6%	92.1%	86.7%
Ratio 65+	97.1%	381.8%	95.5%	88.8%

Source: Author's calculations based on ASEC 2011 and ATUS 2010 data.

3.2. Matching: ASEC 2011 and SCF 2010

For the matching process between the ASEC 2011 and SCF 2010, five strata variables, namely *income categories, home ownership, family type, and race and age of the householder (head of household),* are used to create 360 matching cells. Given the availability of information from both surveys within each cell, and the requirements imposed for consistent estimation of the propensity scores via logit models, we end up with 220 cells in the first round, which represent about 97% of the whole sample.[9]

A dummy variable indicating whether the observation corresponds to the donor or the recipient survey is used as the dependent variable. In addition to the strata variables, a set of variables including: dummies for zero income, zero wage income, dummies for other sources of income, age (and its square) of the householder, education attainment, occupation category and number of people in household, are included in the model specification. Standardized indexes for income and wage income are also included. The logit models and propensity scores are estimated using all information within broader cells, but the matching is elaborated only across observations left unmatched from previous rounds. For subsequent matching rounds, broader matching cells are defined accordingly, keeping the logit specifications consistent across all models, and including the omitted strata variable in the specification

Turning to the results of the match performance, Table 9 shows the distribution of the matched records by matching round. As expected, a large share of the matches (84.6%) occurs on the first round, when the highest level of quality match is ensured. While in the first round the match ratio is lower than in the previous case (ATUS-ASEC), it is still sufficiently large to obtain good matching quality in terms of the strata variables. Only 0.3% of the weighted sample is left unmatched after all matching rounds. These unmatched observations are composed of middle to high income, renter households, with mostly non-elder and predominately Hispanic or White householder. This should not bias the distributional statistics of the transferred information in the aggregate.

[9] For each cell, a minimum of 10 observations from both surveys are require to proceed with the estimation of the propensity score. At the same time, in cases where the logit model indicates perfect prediction of outcomes, the respective observations are excluded for the calculation of the propensity scores.

Table 9 Distribution of Matched Records by Matching Round

Matching Round	Records Matched	Percent	Cumulative Percent
1	100,428,558	84.6	84.6
2	2,469,070	2.1	86.7
3	5,132,203	4.3	91.0
4	5,217,029	4.4	95.4
5	253,425	0.2	95.6
6	560,166	0.5	96.1
7	43,655	0.0	96.1
8	676,621	0.6	96.7
9	1,530,290	1.3	98.0
10	431,460	0.4	98.4
11	379,884	0.3	98.7
12	255,504	0.2	98.9
14	395,453	0.3	99.2
16	545,097	0.5	99.7
17	364,201	0.3	100
Total	**118,682,616**		

Source: Author's calculations based on ASEC 2011 and SCF 2010 matched data.

Table 10 provides a better look at the match quality, comparing some distributional statistics on household's net worth of assets and liabilities. Table 10 also presents some statistics on individual assets and debts categories.[10] The upper percentiles and Gini coefficients are equivalent across both samples. The lower percentiles, however, present a more pronounced difference, with the ASEC presenting lower net worth estimates. This is related to differences in the incidence of homeowners with mortgages shown in Table 3. The differences in the percentiles are also replicated when looking at the percentile ratios. The means and medians show a fair level of equivalence between both surveys for the disaggregated components. The largest difference corresponds to Asset3 (Liquid Assets) showing and average difference of 4% or about $6,600.

[10] Assets are classified in Gross value of housing-asset1, Value of real state and Unincorporated Businesses-Asset2, Liquid assets (checking, saving, cash, etc.)-Asset3, total directly-held mutual funds-Asset4, individual retirement accounts and thrift-type plans-Assets5. Similarly, debts are classified in Housing debt-Debt1 and other debt-Debt2.

Table 10 Matching Quality: Summary Statistics

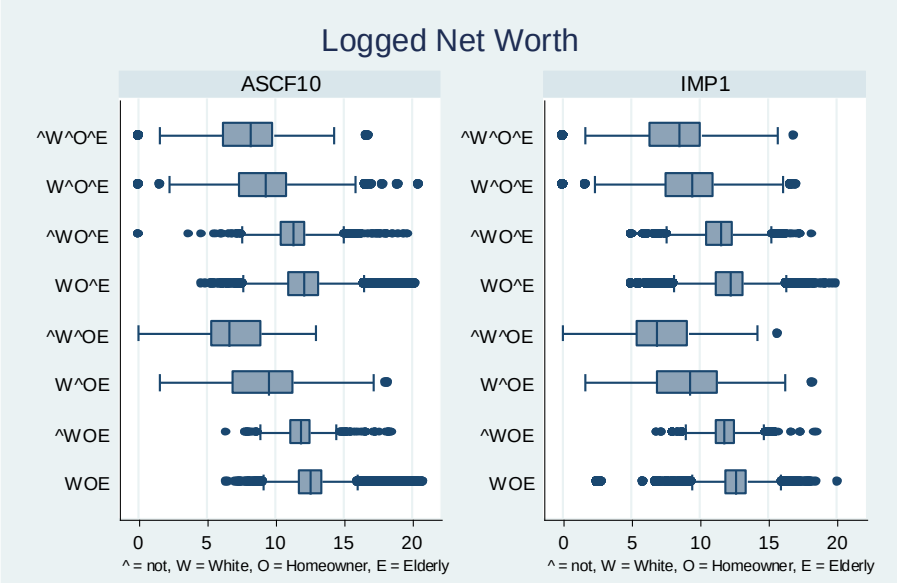
	SCF2010	ASEC 2011	Ratio ASEC/SCF
Distributional Statistics (Net worth)			
p10	-$10,650	-$12,825	120%
p25	$400	$200	50%
p50	$59,430	$62,691	105%
p75	$276,255	283,154	102%
p90	$899,005	$919,503	102%
p90/p50	15	15	97%
p75/p25	691	1,416	205%
p75/p50	5	5	97%
p50/p25	149	313	211%
Gini	0.87	0.87	100%
Summary Statistics			
Average Asset1	$175,555	$172,896	98%
Average Asset2	$163,797	$157,161	96%
Average Asset3	$44,083	$43,405	98%
Average Asset4	$91,974	$91,725	100%
Average Asset5	$78,500	$78,096	99%
Average Debt1	$72,405	$70,963	98%
Average Debt2	$14,701	$14,688	100%
Average Net worth	$466,809	$457,631	98%
Median Asset1	$100,000	$100,000	100%
Median Asset2	$0	$0	
Median Asset3	$4,150	$4,400	106%
Median Asset4	$0	$0	
Median Asset5	$0	$0	
Median Debt1	$0	$0	
Median Debt2	$2,400	$3,000	125%
Median Net worth	$59,430	$6,2691	105%

Note: Assets are classified in gross value of housing-asset1, value of real state and unincorporated businesses-Asset2, liquid assets-Asset3, total mutual funds-Asset4, individual retirement accounts and thrift-type plans-Assets5. Similarly, debts are classified in housing debt-Debt1 and other debt-Debt2

Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

Figure 4 presents a visual representation of the distribution of logged household net worth using three of the strata variables: race, homeownership and age. The figure shows that for most cases the distribution of the logged net worth is equivalent in both surveys. There are, however, some differences in the distributions regarding extreme values (outliers) among some groups, like households with white elderly homeowners (WOE), nonwhite elderly homeowners (^WOE) or white non elderly and non-homeowners (W^O^E). While extreme values might not affect statistics like medians and percentiles, they might create problems when analyzing information at the means for more detailed subgroups.

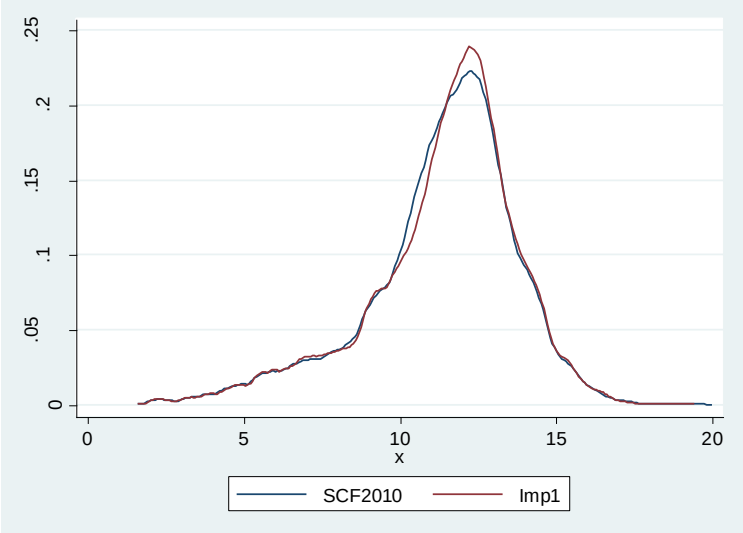
Figure 4 Distritubtion of Logged Net worth, by Survey



Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

Figure 5 provides an alternative to compare the distribution of logged net worth between both the donor and the imputed sample. The close superposition between the kernel densities for both suggests that, as indicated before, the moments of the distributions of the imputed and donor samples are highly comparable in the aggregate. A closer look at Figure 5, however, still indicates that the presence of outliers might affect estimation of relevant means for specific groups. Overall, there is a difference of only $4,948 between the mean imputed and donor Net worth, and no differences when comparing medians.

Figure 5 Kernel density of Logged Net Worth, by Survey



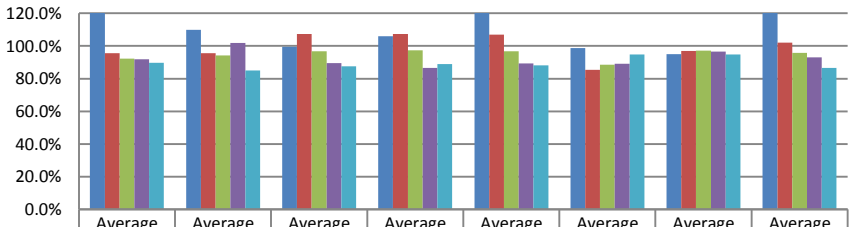
Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

For a detailed review of the performance of the matching, figures 6 and 7 show the ratios of assets and debts values between the imputed data (ASEC) and the donor data (SCF), across the five strata variables used for the matching. Table 11 also presents information on the means and medians gaps of the net worth of the households with respect to the strata characteristics. The first strata variable to be analyzed corresponds to the household income. After the matching, the average values of Asset1, Asset5 and net worth are overstated (up to 21.5%) in the recipient dataset among household in the lowest income group. This implies a difference of a little more than $11,000 for Asset1 or $15,000 for net worth. In contrast, with a few exceptions, all other assets/debts are understated in the imputed dataset by almost 7% in average, with the richest households having the largest bias (14% or 236,000$ lower Net worth). In all cases, Debt1 and Debt2 are understated for all income groups, with a bias of less than 15%.

With respect to homeownership, the results show a good balance in average, with net worth differences from ranging from $4,000 to $80,000. The groups with the largest imbalances correspond to home owners without mortgage, for which Stocks and Bonds (Asset3) is understated by almost 16% and Other Debt (Debt2) is overstated by 34%, and home owners with mortgage, for which Mortgage Debt (Debt1) is overstated by about 16%. In terms of family type, while households with couples and single women have well balanced statistics, liquid assets in single male households are overstated by 40% (Asset2), and net worth overstated in about 22% or $56,000 in average (Table 11).

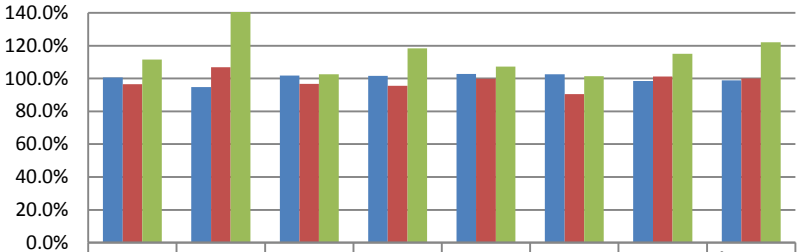
Figure 6 Ratio of Mean Household Assets and Liabilities, by Strata Variables

Houshold Income



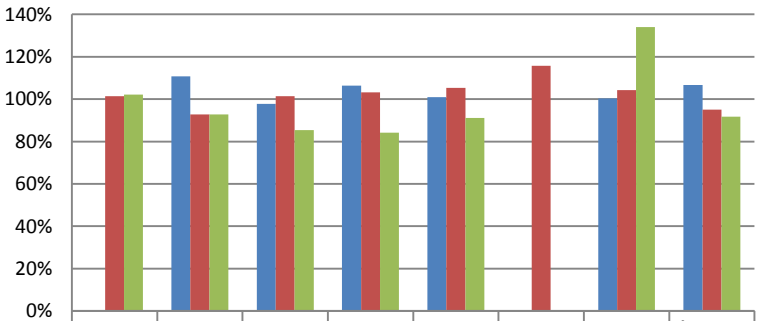
	Average Asset1	Average Asset2	Average Asset3	Average Asset4	Average Asset5	Average Debt1	Average Debt2	Average Networth
ratio lt $20k	121.5%	109.8%	99.4%	106.0%	120.3%	98.7%	94.9%	120.6%
ratio $20-50k	95.5%	95.6%	107.3%	107.3%	106.8%	85.3%	96.8%	102.1%
ratio $50-75k	92.3%	94.1%	96.7%	97.3%	96.7%	88.5%	97.1%	95.7%
ratio $75-100k	91.9%	101.9%	89.6%	86.6%	89.4%	89.2%	96.6%	93.1%
ratio gt $100k	89.7%	85.0%	87.5%	88.9%	88.1%	94.8%	94.7%	86.6%

Family Type



	Average Asset1	Average Asset2	Average Asset3	Average Asset4	Average Asset5	Average Debt1	Average Debt2	Average Networth
ratio Couple	100.7%	94.8%	101.8%	101.6%	102.8%	102.7%	98.5%	98.9%
ratio Single Female	96.5%	106.8%	96.8%	95.6%	99.9%	90.6%	101.3%	100.0%
ratio Siingle Male	111.6%	142.1%	102.6%	118.5%	107.3%	101.5%	115.0%	122.1%

Home Ownership

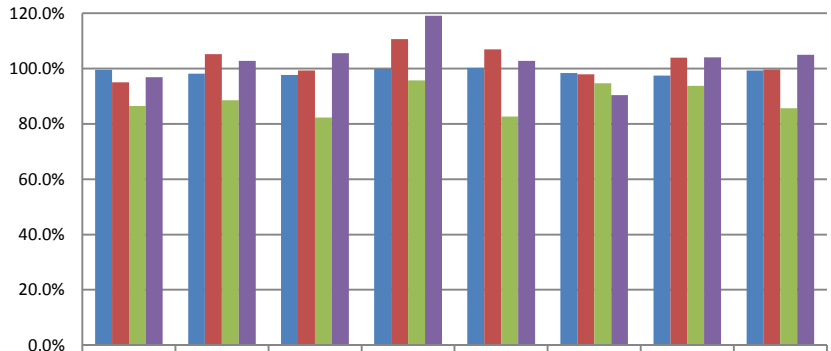


	Average Asset1	Average Asset2	Average Asset3	Average Asset4	Average Asset5	Average Debt1	Average Debt2	Average Networth
ratio Renter		110.8%	97.7%	106.4%	100.9%		100.3%	106.6%
ratio Owner w/mort	101.4%	92.7%	101.3%	103.2%	105.3%	115.7%	104.3%	95.0%
ratio Owner wo/mort	102.1%	92.8%	85.4%	84.2%	91.2%		134.0%	91.8%

Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

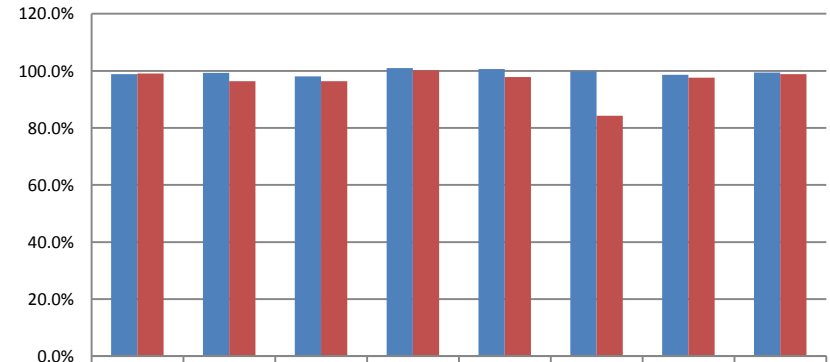
Figure 7 Ratio of Mean Household Assets and Liabilities, by Strata Variables

Race



	Average Asset1	Average Asset2	Average Asset3	Average Asset4	Average Asset5	Average Debt1	Average Debt2	Average Networth
ratio white	99.5%	98.2%	97.6%	99.9%	100.2%	98.4%	97.5%	99.3%
ratio black	95.0%	105.2%	99.3%	110.7%	107.0%	97.9%	103.9%	99.7%
ratio other	86.4%	88.5%	82.3%	95.7%	82.6%	94.6%	93.8%	85.7%
ratio hispanic	96.8%	102.8%	105.6%	119.1%	102.8%	90.4%	104.0%	105.0%

Age group



	Average Asset1	Average Asset2	Average Asset3	Average Asset4	Average Asset5	Average Debt1	Average Debt2	Average Networth
ratio NonElder	98.8%	99.3%	98.0%	101.0%	100.6%	99.8%	98.6%	99.4%
ratio Elder	99.1%	96.4%	96.4%	100.2%	97.8%	84.2%	97.6%	98.8%

Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

When considering race, while the balance statistics show that information corresponding to households with white, black and Hispanic householders is well balanced, the imputed sample consistently understates the assets/debts holdings from other race households by almost 18%. In terms of net worth alone, "Other-races" net assets are understated in just over 18% which imply almost $90,000 difference. The medians gap show a much smaller absolute difference ($20,000), suggesting that the large differences in the means are driven by outliers. Finally, in terms of age groups, the statistics show that the imputed data is well balanced for most of the asset/debt components, except for mortgage debt (Debt1). The statistics show that the imputed data understates elderly household debt in about 16%. This happens because the share of elderly households with mortgage debt is lower in the ASEC survey compared to the corresponding share in the SCF.[11]

Table 11 Mean and Median Net Worth by Strata Variables

	Averages			**Median**		
	Donor	**IMP1**		**Donor**	**IMP1**	
Total	$466,809	$457,631	98.0%	$59,430	$62,691	105.5%
Home Ownership						
Renter	$47,400	$50,534	106.6%	$55	$70	127.3%
Owner Mortg	$541,521	$514,563	95.0%	$110,105	$106,457	96.7%
Owner wo/ Mortg	$973,132	$893,156	91.8%	$263,305	$245,600	93.3%
Income Group						
lt $20k	$73,136	$88,181	120.6%	$1,000	$2,000	200.0%
$20-50k	$170,974	$174,527	102.1%	$30,805	$28,586	92.8%
$50-75k	$244,142	$233,733	95.7%	$76,250	$70,770	92.8%
$75-100k	$302,437	$281,466	93.1%	$117,810	$106,350	90.3%
gt $100k	$1,769,339	$1,532,811	86.6%	$520,000	$398,370	76.6%
Age						
NonElder	$393,449	$389,430	99.0%	$35,000	$34,782	99.4%
Elder	$723,441	$700,001	96.8%	$194,450	$191,801	98.6%
Family type						
Couple	$653,716	$646,596	98.9%	$102,490	$114,697	111.9%
Single Female	$181,994	$182,022	100.0%	$21,620	$19,605	90.7%
Single Male	$252,558	$308,441	122.1%	$28,500	$28,909	101.4%
Ethnicity						
White	$596,808	$588,376	98.6%	$100,900	$110,228	109.2%
Black	$86,188	$80,377	93.3%	$6,050	$03,201	52.9%
Other	$489,367	$399,134	81.6%	$59,105	$38,258	64.7%
Hispanic	$90,887	$96,091	105.7%	$2,900	$2,900	100.0%

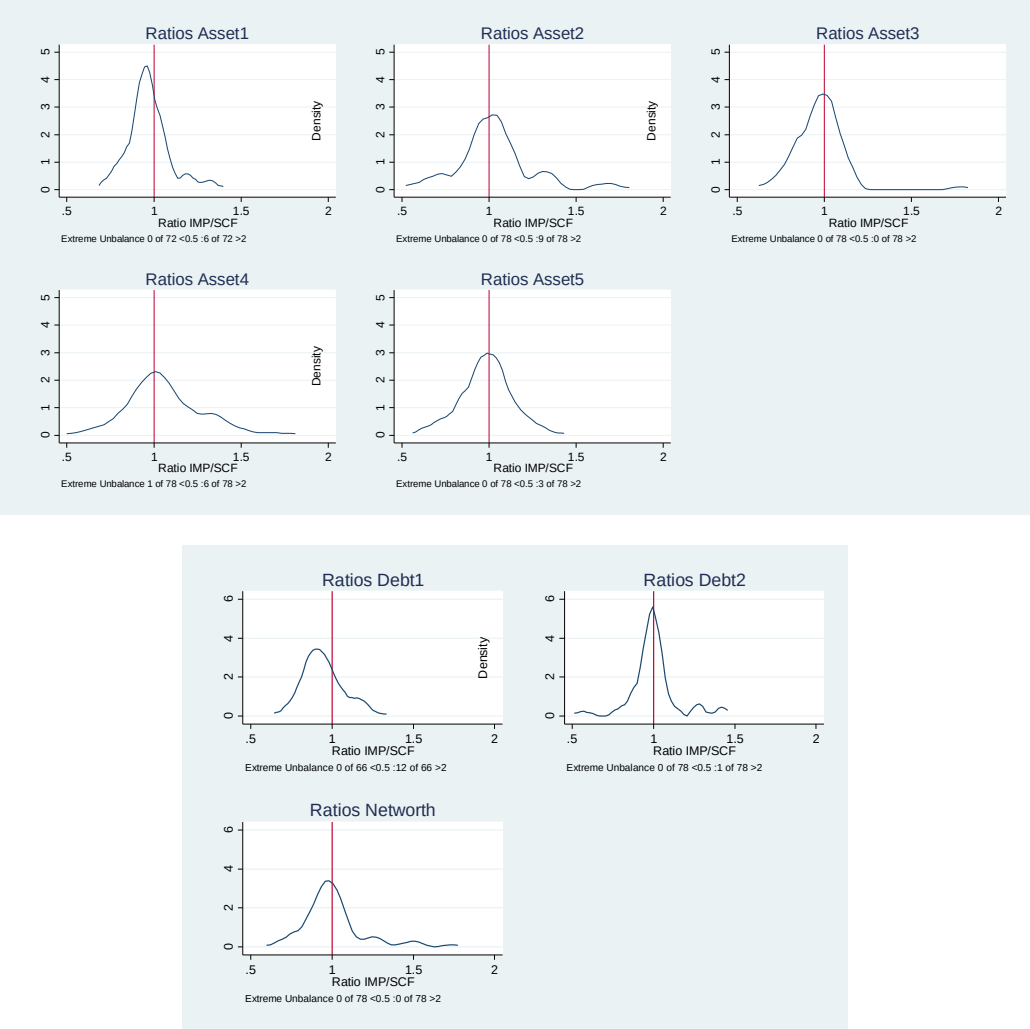
Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

[11] While ignoring mortgage status as part of the strata variables improves the overall balance of item Debt1, it also assigns additional debts to households that should have no mortgage debt.

To analyze how the matching performs for more detailed cells, the mean ratios between samples for all assets and debts are calculated for different combinations of the strata variables.[12] Figure 8 plots the densities corresponding to the mean ratios for selected combinations of the strata variables. As can be seen, for most of the cases, the distributions of the mean ratios are highly concentrated around 1, indicating that, in average, there is good balance between both surveys. As the figure also indicates, for some of the ratios, some large imbalances can be observed (ratios above 2). These types of large imbalances for narrower cells are expected as the SCF also collects information for high income families, which might appear as large outliers. While for most variables, the ratios distributions indicate a good balance, the ones corresponding to Debt1 suggest that the imputed data tends to understate the value of Debt1 (12%).[13] Similarly, house assets (Asset1) also tend to be understated in about 5%.

[12] The cell combinations include: race-homeownership, race-age group, race-family type and race-income group.
[13] It should be noticed that the level of bias is larger if the information were not to be redistributed.

Figure 8 Kernel densities Ratios of Mean Household Assets and Liabilities



Source: Author's calculations based on ASEC 2011 and SCF 2010 data.

4. CONCLUSIONS

Overall, the ATUS and ASEC data are well aligned, with the some imbalances with respect to labor force status. The matching quality is good, with some limitations. There is a strong balance across the individual strata variables, showing good balance for aggregate measures (household production) for most of the variables analyzed. The results across the individual matching cells and other variables, however, show less balance.

On the one hand, the imputed information on the hours allocated to care activities shows important (relative) imbalances across many matching cells. The absolute differences, however, are small and should not create a large bias. On the other hand, information across other variables, such as education, household income, and particularly age, show important balance problems. The imputed dataset over-states household production of people with less than a high school education, understates for those with tertiary education, as well as for people in poor households. Across age, while the aggregate results are balanced, the individual components show large over and under estimations for different age groups.

With respect to the SCF and ASEC, the data is also well aligned, with the exception of house mortgage holding, and small difference in the proportions of the breakdown by sex of the householder. The results regarding the quality of the match are mixed. While the overall results show good balance between the imputed and donor surveys, with small under-estimations of some items, analyzing the results across the strata variables show relatively large imbalances (up to 20%) for a relatively small subset of strata variables. As we would expect, larger imbalances are observed for narrower groupings. The data shows some underestimation of mortgage debt, probably caused by the differences on the alignment of household property (see table 3). Given that the SCF collects information from high income households, it is possible that the information transferred from these observations has a strong influence on the cell specific statistics. These results imply that careful consideration must be taken when making statistical inferences from certain populations. One can make inferences for the aggregate population, but attempting a similar analysis using two or more variables at the same time may carry too much bias to be informative.

REFERENCES

Kennickell, A.B. (2000) "Revisions to the Variance Estimation Procedure for the SCF", Washington:Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/econresdata/scf/files/variance.pdf

Kennickell, A.B. and Woodburn, R.L. (1999) "Consistent Weight Design for the 1989,1992, and 1995 SCFs, and the Distribution of Wealth," *Review of Income and Wealth,* 45(2): 193-215

Kum, H., & Masterson, T. N. (2010). Statistical matching using propensity scores: Theory and application to the analysis of the distribution of income and wealth. *Journal of Economic and Social Measurement, 35*(3), 177-196. doi: 10.3233/JEM-2010-0332

Wolff, E. N., & Zacharias, A. (2003). The Levy Institute Measure of Economic Well-Being. *The Levy Economics Institute Working Papers, 372*.