

Reconciling LFPR and income measures

Madison informal notes

June 28, 2023

Issue: My LFPRs by income for groups defined by their **family** income differed significantly from Shigeru's LFPRs by income groups defined by their **household** income for the \$25,000< group. There are some legitimate reasons for the disparity, which I cover toward the end of this document. However, investigation revealed a group of people in single-family households with family income < \$25,000 and household income \geq \$25,000, which is nonsensical and prompted suspicion. Further investigation exposed a slew of issues around the use of my constructed variable, `unlog_ftotval`, and the way we include in our imputation model ASEC observations where individuals have zero or negative values for family or household income. I describe the problem and my solution. Then I show recomputed LFPRs by family income using the corrected family income variable. Finally, I break apart the difference between my calculations and Shigeru's.

Define variables used:

- `htotval`: CPS given household income
- `ftotval`: CPS given family income. This includes cases where `ftotval` ≤ 0 .
- `unlog_ftotval`: $e^{\ln(ftotval)}$. This is the variable I use for family income. It takes missing values for people with `ftotval` ≤ 0 , unless an intervention is made.
- `h_tot_income_calc`: (`htotval` standard items)¹ + energy assistance + food stamps + COVID economic impact payments. Takes value zero if its component parts are less than zero²

1 `ftotval` and `unlog_ftotval`

Problem: `unlog_ftotval` is the variable we must use for family income after running the log-linear imputation model. The left-hand side variable is `ln(ftotval)` so only those with `ftotval` > 0 can be taken in and used to predict. The predicted values generated by the model are `ln(predicted ftotval)` so I must exponentiate those values again to get predicted family income in dollars again. Because the prediction model is log-linear, all predicted values will also be > 0 . This means our model will systematically over-predict family income because it is incapable of predicting a zero for family income but clearly we do have those cases in the ASEC data. Inability to incorporate those cases likely compromises the accuracy of the imputation.

¹See Shigeru's March memo for thorough definition.

²replace `h_main_income` = 0 if `h_main_income` < 0
replace `h_other_income` = 0 if `h_other_income` < 0
gen `h_tot_income_calc` = `h_main_income` + `h_other_income` + `h_ben`

Table 1: Frequency of people with $FTOTVAL \leq \$0$ in ASEC

	Year		
	2020	2021	2022
ftotval==0	1,918	2,290	2,161
ftotval<0	30	73	46

The question here is how people with negative or no family income ought to be incorporated. What I was doing was letting the people with $ftotval \leq 0$ take missing values for $\ln_ftotval$, ultimately resulting in them having missing values for unlog_ftotval . This meant that when I was computing LFPRs by family income using unlog_ftotval , I was excluding those people. This is a consequential misstep and one of the reasons my LFPR computation was too high for the below-25k group. The following table illustrates how different the estimates are if we rectify the exclusion. The top row is the group for whom I had previously tabulated and the bottom is the corrected group.

Table 2: Comparing LFPR computations

	Year		
	2020	2021	2022
$0 < ftotval < 25k$	36.4	34.2	34.0
$ftotval < 25k$	34.3	32.1	32.2

My proposed solution is to replace $\ln(ftotval)$ with zero if $ftotval \leq 0$. This is in-keeping with Shigeru’s practice of replacing all negative income values with zero in constructing $h_tot_income_calc$. Doing this allows those observations to be part of the information set that the regression uses to predict income values, which also rectifies some of the issues of over-prediction at the lowest end of the income distribution.

*To flag a potential issue, we know from prior investigation that negative income values are reflective of a business owner taking losses, so it’s possible that people with negative incomes shouldn’t really be lumped into the lowest-income category.

All computations for the rest of this paper will be done within the universe of $ftotval < 25,000.00$. When I say family income, I am referring to $ftotval$. Bear in mind that $\text{unlog_ftotval} == ftotval$ for all except those with $ftotval \leq 0$; in those cases, $\text{unlog_ftotval} == 1$ because I replaced $\ln(ftotval)$ with 0 instead of a missing value. Since I am using fixed threshold income categories here *AND* I have made aforementioned replacement, there is no difference in using $ftotval$ or unlog_ftotval to define the family income categories for the ASEC observations. Cases with $ftotval \leq 0$ end up in the bottom bin either way. What’s important is that those cases are now included in the bottom income category ($\$0$ - $\$25,000$) instead of being excluded from computations. In all computations, I have used the March Supplement frequency weights, $marsupwt$.

2 Families and Households

Families are different from households. Households may be single-family or multi-family. Everyone with a household income under \$25k also has a family income under \$25k because in the CPS, families are defined as being a level below household. Household income is equivalent to the sum of the family incomes for each

family in the household³. This allows for the existence of cases where two families each have family incomes below 25k which sum to a household income greater than 25k.

Consider four cases in the universe of individuals with family income less than \$25,000:

1. In single-family household, family income < \$25,000, household income < \$25,000
2. In single-family household, family income < \$25,000, household income \geq \$25,000
3. In multi-family household, family income < \$25,000, household income < \$25,000
4. In multi-family household, family income < \$25,000, household income \geq \$25,000

Household income being the sum of all family incomes for each family in a household means that there are no instances of the second type⁴. Previously, when I was doing the computations in universe `unlog_ftotval<$25,000` and I had left missing values for everybody `ftotval \leq 0`, many of these individuals appeared in that group simply because of the high-coded missing value.

Table 3: Frequencies - Family income<\$25,000 Universe

Persons	Year		
	2020	2021	2022
Single-family, hh<25k	12,883	14,559	13,224
Single-family, hh \geq 25k	1	0	0
Multi-family, hh<25k	1,216	1,484	1,266
Multi-family, hh \geq 25k	3,738	4,166	3,628
Total, hh<25k	14,099	16,043	14,491
Overall total	17,838	20,209	18,118

My computations of labor force participation were based on family income only, so my below-25k group contains cases 1, 3, and 4 (Overall total). Shigeru's computations of labor force participation were based on household income so his below-25k group contains people from cases 1 and 3 (Total, hh<25k).

³In years 2020, 2021, and 2022, there are 11, 13, and 17 observations, respectively, where this is not true. This makes for a total of 41 out of 367,781 observations.

⁴The singular individual in 2020 in a single family household with family income less than household income puzzles me. Household income is \$239,860 and family income is \$9,600, yet there is only one individual in the household. She is a widower in her 70s, listed as the householder of her house so I don't think it's a group-home situation. She is listed as being NILF - retired but has a listed wage/salary income of \$220,000. My inclination is to ignore this one observation.

Table 4: LFPRs - Family income<\$25,000 Universe

% as of March	Year		
	2020	2021	2022
Single-family, hh<25k	28.0	26.1	25.6
Single-family, hh≥25k	-	-	-
Multi-family, hh<25k	35.9	34.8	41.1
Multi-family, hh≥25k	55.1	52.2	52.3
hh<25k	28.7	26.9	27.1
Overall total	34.3	32.1	32.2

Notes: Sample is all participants 16+ from the March Supplements with FTOTVAL<\$25k. Weights used are marsupwt.

Notice that the overall LFPR for that group (Total, hh<25k) is not exactly as Shigeru had calculated in March. We will see one reason for this in a moment.

Notice also how much higher the LFPR is for the group that is present in my calculations but not Shigeru's. It makes sense that my number was higher. I will confirm this at the end by showing a sum of weighted averages.

3 HTOTVAL vs. h_tot_income_calc

Shigeru's household income measure was not simply the one given by CPS (HTOTVAL). Constructing variable h_tot_income.ASEC_calc, he showed exactly which income line items are encompassed in htotval (see memo from March). To that sum, he added in energy assistance, food stamps, and COVID economic impact payments to create variable h_tot_income_calc. He defines the income groups using *this* measure of household income. This measure takes larger values than HTOTVAL, especially in years 2021 and 2022 because that's when the EIPs are being paid out. This means that there will be more people in those years who fall into the case of family income<25k and HH income>25k during when we use these numbers for HH income. This is shown here:

Table 5: Frequencies of people with FTOTVAL<\$25k

Persons	Year		
	2020	2021	2022
h_tot_income_calc ≥\$25k	4,174	7,503	6,692
h_tot_income_calc <\$25k	13,664	12,706	11,426
Total	17,838	20,209	18,118

Notes: Sample is all participants 16+ from the March Supplements with FTOTVAL<\$25k

Table 6: Mean LFPR of people with FTOTVAL<\$25k

% as of March	Year		
	2020	2021	2022
h_tot_income.calc \geq \$25k	54.1	45.0	46.6
h_tot_income.calc <\$25k	28.2	24.7	23.9
Total	34.3	32.1	32.2

Notes: Sample is all participants 16+ from the March Supplements with FTOTVAL<\$25k

% as of March	2020	2021	2022
Shigeru's March numbers	27.8	24.4	24.1

The second row corresponds to the group for whom Shigeru computed labor force participation. As we see, my these values are very close to those he computed in March. I still do not have a suitable answer for why they are not equal.

Here are the proportions each case constitutes of the overall subsample (family income<25k):

Table 7: Proportion of subsample

Persons	Year		
	2020	2021	2022
HH income \geq \$25k	0.23	0.37	0.37
HH income <\$25k	0.77	0.63	0.63
Together	1.00	1.00	1.00

Notes: Sample is all participants 16+ from the March Supplements with FTOTVAL<\$25k

Multiplying each LFPR by its respective group's proportion and then summing yields the same values as in the 'Total' row of Table 6.