# SCF Calibration Targets

This document describes how the creation of calibration targets from the Survey of Consumer Finance. For now I have used the SCF from 2004.

**Version 1.3**: Added some stuff about calibrating initial log(permanent income) based on SCF data.

## Sample selection

We apply the following sample selection criteria:

* Only include households where the head of the households is between 25 and 62 (working age).
  + 1,188 households dropped
* Exclude households where **norminc** < 0.
  + 0 observations dropped
* Exclude households where **norminc** is in the bottom 5 % of the distribution.
  + 134 observations dropped, minimum value of norminc we keep is $16,708

## Permanent income

For permanent income I use the SCF variable **norminc**.

## Liquid wealth

I define liquid assets using the same definition as Kaplan et al. (2014a):   
 “In the U.S. SCF, our definition of liquid assets consists of checking, saving, money market, and call accounts as well as directly held mutual funds, stocks, corporate bonds, and government bonds.”

In addition, I use the same adjustment for cash holdings (which are not in the SCF) as described in Kaplan et al. (2014b), appendix B1.

The liquid assets variable is then defined as follows from SCF variables:   
liquid assets = 1.05 \* (checking + saving + mma + call) + nmmf + stocks + bond

I also define liquid debt as in Kaplan et al. (2014a):   
“We define liquid debt in the U.S. SCF as the sum of all credit card balances that accrue interest, after the most recent payment.”

But as in Kaplan et al. (2014b), appendix B1 I only include balances for those who only “Sometimes” or “Hardly ever” repay their balance in full. That answer is captured in the SCF variable X432 which is in the full public data set (but not in the summary extract public data). After that adjustment, we have:  
liquid debt = ccbal

After defining liquid debt, we have:

Liquid wealth = liquid assets – liquid debt

## Imputation and implicates

For each household (or more precisely, “principal economic unit”) in the SCF the record for the household is repeated 5 times. This has to do with both the way the SCF handles missing data and with masking the identity of a respondent. From Hanna et al. (2018): “The SCF actually calculates five different estimates for a missing value and thus, for each household, provides five complete datasets. Each of these datasets is called an implicate.”

Whether a researcher handles this by simply using one of the implicates, by combining them all into one dataset or by some other method, seems important if the aim is to run regressions and do various tests for statistical significance. For our purposes though, I don’t think it is very important. What I have done is to simply **average all variables across the five implicates**. This includes averaging the weights associated with the household. Let me know if you think this is an issue.

## Weights

Having averaged the weights, if I sum all the weights and multiply by 5 (since I am using each household once instead of 5 times), I get that the total number of households in the US is about 112 million which seems about right.

## Outliers

I have dropped outliers as follows:

1. Drop anyone with negative liquid wealth. Hence, the Lorenz curves never dip below zero.

🡪 drop 476 observations

## Education

The SCF classifies education into four groups, but I combine two of them to get the three referred to above. After dropping outliers, the population is divided into these groups as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Education completed: | No high school | High school/some college | College |
| Percent of population: | 9.3 | 52.7 | 38.0 |
| Percent of total wealth | 0.8 | 17.9 | 81.2 |

## Calibration targets 1: Average liquid wealth / permanent income by education group

To calculate these targets for each education group I do as follows:

1. Rescale the weights so that I have a set of weights that sum to 1 for each education group.
2. Calculate liquid wealth / permanent income for each household.
3. Calculate the weighted average of liquid wealth / permanent income for each education group using the rescaled weights.

I also calculate total liquid wealth / total permanent income for each education group, and now I also calculate the weighted median of the individual liquid wealth / permanent income ratios.

Results:

|  |  |  |  |
| --- | --- | --- | --- |
| Education completed: | No high school | High school/some college | College |
| Weighted avg. LW/PI | 0.157 | 0.477 | 1.11 |
| Total LW / total PI | 0.281 | 0.596 | 1.62 |
| Weighted median LW/PI | 0.0116 | 0.0755 | 0.282 |

Note: numbers are annual, to convert to quarterly multiply by 4 (divide denominator by 4)

|  |  |  |  |
| --- | --- | --- | --- |
| Education completed: | No high school | High school/some college | College |
| Weighted median LW/PI annual | 0.0464 | 0.302 | 1.128 |
| Weighted median LW/PI annual at age 62 | 0.0721 | 1.82 | 7.28 |

## Calibration targets 2: Lorenz curves and percentiles

To calculate Lorenz curves I calculate the weighted sum of liquid wealth for all households and sort them by liquid wealth. Then I plot the cumulative sum of the fraction of weighted liquid wealth against the sum of the weights. Both are multiplied by 100.

The Lorenz curves I get as well as the [20,40,60,80] percentiles for all households are as follows:



[20,40,60,80]-percentiles of weighted liquid wealth: [0.029, 0.354, 1.84, 7.42]

## Calibration targets 3: Lorenz curves and percentiles, by education group

Finally, to calculate Lorenz curves I calculate the weighted sum of liquid wealth for each education group. Then I sort households by liquid wealth within each group and plot the cumulative sum of the fraction of weighted liquid wealth against the sum of the weights. Both are multiplied by 100.

The Lorenz curves I get as well as the [20,40,60,80] percentiles for each education group are as follows:



|  |  |  |  |
| --- | --- | --- | --- |
| Education completed: | No high school | High school/some college | College |
| [20,40,60,80]-percentiles of weighted liquid wealth: | [0, 0.01, 0.60, 3.58] | [0.06, 0.63, 2.98, 11.6] | [0.15, 0.92, 3.27, 10.3] |

## Calibrating distribution of initial permanent income

We took the numbers for the distribution of initial permanent income from the pandemic paper, but in that paper it is not clear where those numbers come from. An alternative is to take them from numbers for 25-year olds in the SCF.

|  |  |  |  |
| --- | --- | --- | --- |
| Education completed: | No high school | High school/some college | College |
| Weighted avg of permInc at age 25 (quarterly) | $6,421 | $12,063 | $16,739 |
| Weighted avg of log(permInc) at age 25 (q) | 8.7272 | 9.3125 | 9.5849 |
| Exp of # above | $6,168 | $11,076 | $14,544 |
| Std. dev. of weighted log(permInc) at age 25 (quarterly) | 0.32 | 0.42 | 0.53 |

The standard deviation of the weighted log(permInc) (quarterly) for the whole population is 0.52.

Histograms of log(initial PI) for each education group look as follows (Note: histogram is unweighted, not sure how to construct a weighted histogram in Stata):



## Thoughts/Conclusions

In the calibration of unemployment by education group, we just take a simple average of the unemployment of those with highschool degrees, some college and Associate’s degrees. No attempt is made to use SCF data to weight these three groups.

## References

* Kaplan, Violante and Weidner (2014a), “The Wealthy Hand-to-mouth”, *Brookings Papers on Economic Activity*, vol. 45, issue 1 (Spring), 77-153
* Kaplan and Violante (2014b), “A Model of the Consumption Response to Fiscal Stimulus Payments”, *Econometrica*, Vol. 82, No. 4 (July), 1199–1239
* Hanna, Kim and Lindamood (2018), “Behind the Numbers: Understanding the Survey of Consumer Finances”, Journal of Financial Counseling and Planning, Vol 29, Issue 2