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.

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

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 us 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.
2. Within each education group, drop the top 0.5 percent of liquid wealth holders.
   Note: This does not lead us to drop a lot of observations, but we drop ***a lot*** of liquid wealth. For the three education groups we are left with the following percentages of total liquid wealth:

Education completed:	No high school	High school/some college	College
Percentage of liquid wealth left:	51.0	72.4	70.0

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
Percentage of population:	14.7	52.7	32.6

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.

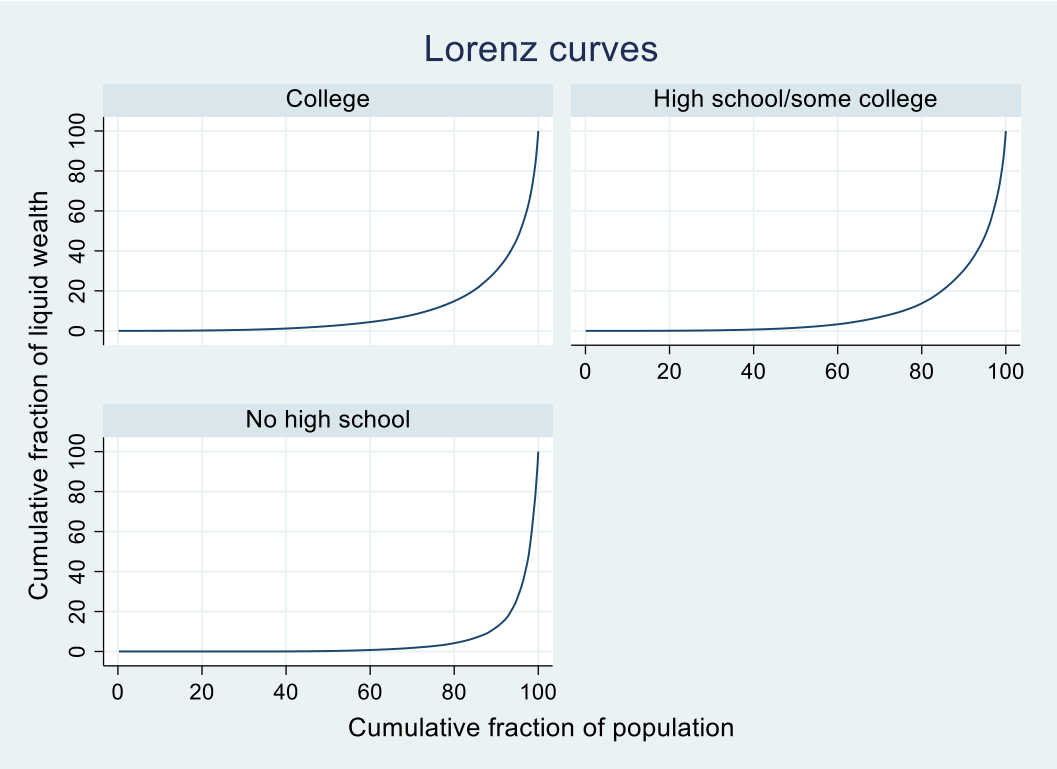
Results:

Education completed:	No high school	High school/some college	College
Weighted average LW/PI	0.432	0.902	1.515

Calibration targets 2: Lorenz curves and percentiles

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, 0.032, 0.73, 4.15]	[0.05, 0.65, 3.27, 13.7]	[0.17, 1.16, 4.39, 14.8]

Thoughts/Conclustions

The sample selection criteria are very important for the final targets that we get. In this version I haven't done any seletion on age (Kaplan et al. (2014a) only include households where the head is between 22 and 79), and I have dropped the top 0.5 percent of liquid wealth rather than using another variable (Kaplan et al. (2014b) drop the top 5 % based on net worth). I also dropped households based on negative liquid wealth (Kaplan et al. (2014a) drop households based on negative income). A discussion of these types of criteria would be very useful.

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