

Problem Set 1: Due Friday, September 19

Note: For this problem set, and all future problem sets, e-mail me your solutions by 3:30 pm of the due date; no need to print out anything.

Problem 1

In this problem, we will re-evaluate some of the results presented in Heathcote, Perri, and Violante (2010). In particular, we will update their CPS-based results using data spanning the Great Recession.

Preliminary Steps: The easiest way to retrieve the data is to download them from the IPUMS website (see <https://cps.ipums.org/cps/>). There, you'll find an easy-to-use website through which you can access the data. Download the March supplements, beginning in 1976 and extending to 2013. For the first part of the sample (up to 2005), feel free to download every other year or every third year if you want to economize on memory. From 2006 to 2013, download every year. The variables you should download are: year; serial; hwtsupp; region; statefip; metro; hhincome; housret; cpi99; month; pernum; wtsupp; relate; age; sex; race; marst; educ; higrade; educ99; empstat; labforce; occ1990; ind1990; wkswork1; hrswork; uhrswork; fullpart; ftotval; inctot; incwage; incbus; incfarm; incss; incwelfr; incgov; incidr; incaloth; incretir; incssi; incdrt; incint; incunemp; incwkcom; incvet; incsurv; incdisab; incdivid; incrent; inceduc; incchild; incalim; incasist; incother; earnweek; incdisa1; incdisa2; inclongj; increti1; increti2; incsurv1; incsurv2; oincbus; wkswork2; srcearn; oincwage; oincfarm; incidr. There is an option for dataset format; choose STATA.

Approximately fifteen minutes after submitting the download request, you should get an e-mail with a link to the dataset. From my website you can also find a *.do* file which will clean the data and construct the main household income variables. The first few questions pertain to some of the details in the code that I have posted.

1. *Top coding* refers to the censoring of certain observations to maintain the confidentiality of survey respondents. In our context the CPS income variables greater than some variable-year-specific threshold are censored.
 - (a) What distributional assumption is made on each income variable to impute the true income of the censored observations?
 - (b) What parameter, according to this distribution, does the coefficient from the $\log y$ vs. $\log v$ regression represent?

- (c) Why is it necessary to run these regressions separately for each year?
 - (d) What is the significance of restricting the sample of this regression to be greater than the "x1" local macro variable?
2. Use one paragraph to describe how you would modify the code if you thought the tails of the income variables were log-normally distributed.
 3. Something we touched on in class: Figures 8 and 9 of the paper use the term "Equivalentized" as a modifier to household earnings.
 - (a) How is equivalentization accomplished in the code that I posted?
 - (b) Do the coefficients in the equivalentization procedure (1 for the first adult, 0.7 for each additional adult, and 0.5 for each additional child) make sense? How would you use the data to check whether these coefficients are appropriate?

The next few questions ask you to produce some figures and tables. When making the figures and tables, make sure to include a short description describing how it was constructed. In the figures, label each axis and data series, using a sensible name (e.g., "Wage Income" and not "incwage").

4. How has equivalentized household income inequality (according to the following measures: labor earnings, net asset income, pre-government income, and pre-tax income) evolved since 2005? Plot the 90-50 and 50-10 ratios of the aforementioned income measures from 1976 to 2013.
5. Define y_{lt} as equivalentized household earnings, y_{kt} as equivalentized household asset income, $y_{k+l,t}$ as the sum of the two, \bar{y}_{it} as the year-t average income (for $i \in \{l, k, k+l\}$), and $s_{kt} \equiv \frac{\bar{y}_{kt}}{\bar{y}_{k+l,t}}$. Note the formula

$$\begin{aligned}
 \text{Variance} \left(\frac{y_{k+l,t}}{\bar{y}_{k+l,t}} \right) &= \underbrace{(s_{kt})^2 \cdot \text{Variance} \left(\frac{y_{kt}}{\bar{y}_{kt}} \right)}_{\text{Term 1: Asset Income}} + \underbrace{(1 - s_{kt})^2 \cdot \text{Variance} \left(\frac{y_{lt}}{\bar{y}_{lt}} \right)}_{\text{Term 2: Labor Earnings}} \\
 &\quad + \underbrace{2 \cdot s_{kt} \cdot (1 - s_{kt}) \cdot \text{Covariance} \left(\frac{y_{lt}}{\bar{y}_{lt}}, \frac{y_{kt}}{\bar{y}_{kt}} \right)}_{\text{Term 3: Covariance btw. Asset and Labor Income}}. \tag{1}
 \end{aligned}$$

Plot the three components of the right-hand side of Equation 1 for the 1976 to 2013 period. What, if anything, does this exercise tell you about the sources of earnings-plus-capital income inequality?

6. David Autor and co-authors¹ argue that part of the increase in 90-50 inequality is due to a reduction in the demand due to "middle-skill" occupations, in particular occupations that are centered around routine tasks, such as clerical work and goods production. To explore this hypothesis, rank occupations (using the occ1990 variable) according to skill (and use the mean hourly wage paid in 1976 to proxy for skill). For each occupation skill percentile, compute and plot the percentage change in a) hours worked and b) hourly wages in the following intervals i) 1976 to 2007, ii) 2007 to 2013, and iii) 1976 to 2013. These figures should, in form, resemble Figure 1 of Autor and Dorn (2013). Describe, in a paragraph, your findings from these figures.

Problem 2

In the second part of the assignment, you will work through one of the following two prompts. The goal of this exercise is for you to get some practice analyzing micro data and relating the patterns you find to the existing literature (in particular why the results are interesting to macroeconomists).

Write up your findings in two-to-four pages, describing your method, your empirical findings, what these empirical findings suggest for the papers we have discussed in class. Present the findings clearly and intelligibly. In particular, all of your tables and figures should include a short description, and all of your figures should be properly labeled.

A. Estimating Home Production Technology

In our discussion of Aguiar, Hurst, and Karabarbounis (2013), we showed how—by looking at the relationship between changes in home production time and market work time—it is possible to estimate the preference elasticity of substitution between non-market and market commodities. This estimate was based on a given assumption regarding home production technology, namely that i) the correlation between home and market technology shocks was 0.7, and that ii) the standard deviation and persistence of the home production technology was the same as that of the market technology. Is it possible to use information on the volatility of states' home production hours and market production hours to estimate the parameters of the technology stochastic processes? If such a method exists, try to implement it. If no such method exists, explain why.

¹See, for example, Acemoglu and Autor (2011) and Autor and Dorn (2013).

B. International Trends in Leisure and Home Production

How has time spent on leisure and home production changed over time, across countries? To address this question, use the The Multinational Time Use Study (here: <http://www.timeuse.org/mtus/>), which tracks time use patterns for 10+ (mainly developed) countries. Are the patterns documented by Aguiar and Hurst consistent with the experiences of other countries (Consider looking at different sub-populations: men vs. women, college educated vs. not, old vs. young, etc... Also, when reporting the evolution of the distribution of leisure, report both the unconditional changes and the changes after adjusting for demographics)? If not, what could account for the observed differences between the US experience and those for other countries? Be careful to make sure that whatever differences you report do not simply reflect differences in the way in which countries' time use surveys have been collected and coded. Check out the MTUS User Guide. For planning purposes, note that the application process, to gain access to this dataset, takes a couple of days to be processed. If you plan on working on this problem, send in the application right away.