Problem Set 1: Due Tuesday, September 8

Note: For this problem set, and all future problem sets, email me your solution by 6:00pm on the due date. No need to print anything out.

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 2014. For the first part of the sample (up to 2006), feel free to download every other year or every third year if you want to economize on memory. From 2007 to 2014, 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 logy vs. logv 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 "Equivalized" as a modifier to household earnings.

   (a) How is equivalization accomplished in the code that I posted?

   (b) Do the coefficients in the equivalization 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 equivalized 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 2014. What has changed since the beginning of the Great Recession?

5. Define y_{lt} as equivalized household earnings, y_{kt} as equivalized 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

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

   (a) Plot the three components of the right-hand side of Equation 1 for the 1976 to 2014 period. What, if anything, does this exercise tell you about the sources of earnings-plus-capital income inequality?

(b) In the paper, we were looking at the distribution of $\log y_{it}$, not y_{it} as is in Equation (1). Why did I omit the "log" in this part of the homework? (Hint: Is there an analogous variance decomposition of $\log \frac{y_{k+l,t}}{\bar{y}_{k+l,t}}$? If not, why not? If yes, what is it?)

(c) Perhaps, though, the time trends of Variance$\left(\log\left(\frac{y_{k+l,t}}{\bar{y}_{k+l,t}}\right)\right)$ will "look similar" to the time trend of

$$(s_{kt})^2 \cdot \text{Variance}\left(\log\left(\frac{y_{kt}}{\bar{y}_{kt}}\right)\right) + (1 - s_{kt})^2 \cdot \text{Variance}\left(\log\left(\frac{y_{lt}}{\bar{y}_{lt}}\right)\right)$$
$$+2 \cdot s_{kt} \cdot (1 - s_{kt}) \cdot \text{Covariance}\left(\log\left(\frac{y_{lt}}{\bar{y}_{lt}}\right), \log\left(\frac{y_{kt}}{\bar{y}_{kt}}\right)\right)$$

Check to see if this is the case.

6. David Autor and co-authors[1] 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 2014, and iii) 1976 to 2014. These figures should, in form, resemble Figure 1 of Autor and Dorn (2013). Describe, in a paragraph, your findings from these figures.

Problem 2

On Wednesday, we'll start discussing Ngai and Pissarides. The subsequent questions consider the draft of the paper given at

http://eprints.lse.ac.uk/4468/1/Structural_Change_in_a_Multi-Sector_Model_of_Growth.pdf

.

I recognize that we haven't gone over this paper, at all, yet. A goal of this exercise is to become familiar with the paper before we discuss it in class.

1. What is the contribution of this paper? What gap in the literature does this paper fill?

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

2. How does the definition of the balanced growth path for Ngai and Pissarides differ from that of Kongsamut et al.? For which paper is a constant real interest rate a necessary condition for a balanced growth path? If the real interest rate is changing over time, *what* is staying fixed?

3. Our goal in this problem and the next is to derive some of the optimality conditions given in page 5. Write out the *Hamiltonian*.[2]

$$\mathcal{H}(t) = e^{-\rho t} v(c_1, ...c_m) + \lambda(t)\left[A_m F(n_m k, n_m) - c_m - (\delta + \nu) k(t)\right] \tag{A}$$

   Here we have already employed the assumptions given in Equation 7. Also, if it makes the math easier for you, free to use the assumption that $F(n_i k_i, n_i) = k_i^{\alpha} n_i$.

   To derive the static optimization condition given in Equation (10), substitute the c_i out of the first term in the right hand side of our Hamiltonian. Then take derivatives with respect to n_i, k_i, n_m and k_m. Using Equation (8), you should be able to derive, now, Equation (10) of the paper.

4. Now, the dynamic efficiency condition. Begin with the intertemporal optimality condition:

$$\frac{\partial \mathcal{H}(t)}{\partial k(t)} = -\dot{\lambda}(t)$$

   The above equation implies that

$$-A_m F_K(n_m k, n_m) + (\delta + \nu) = \frac{\dot{\lambda}(t)}{\lambda(t)} \tag{B}$$

   Now, take first order conditions of Equation A with respect to $c_m(t)$ and differentiate this first order condition with respect to time, to get and expression for $\frac{\dot{v}_{c_m}}{v_{c_m}}$. Substitute out the $\frac{\dot{\lambda}(t)}{\lambda(t)}$ term, using equation B, to arrive at Equation (6) of the paper.

[2] If you are unfamiliar, or need a refresher, http://krebs.vwl.uni-mannheim.de/fileadmin/user_upload/krebs/pdf/Hamiltonian.pdf derives the Hamiltonian function. Todd Keister provides some nice intuition of the Hamiltonian method in Section 2 of http://www.toddkeister.net/pdf/optimal-growth-notes.pdf