Problem Set 4: Due Wednesday, February 15

For this problem set you have a choice: In addition to Problem 3, you should do *either* Problem 1 or Problem 2.

Problem 1

In this problem, we will estimate a many ($\gg 3$) commodity version of the model presented in Herrendorf, Rogerson, and Valentinyi (2013).

Modify the preferences of the representative consumer to have the following utility function:

$$\begin{aligned}
u\left(c_{at}, c_{mt}, c_{st}\right) &= \left[\sum_{i\in\{a,m,s\}} \omega_i^{\frac{1}{\sigma}} \left(c_{it} + \bar{c}_i\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \text{ where} \\
c_{at} &= \left[\sum_{j\in I_a} \left(\omega_j^a\right)^{\frac{1}{\sigma_a}} \left(c_{jt} + \bar{c}_j\right)^{\frac{\sigma_a-1}{\sigma_a}}\right]^{\frac{\sigma_a}{\sigma_a-1}}, \\
c_{mt} &= \left[\sum_{j\in I_m} \left(\omega_j^m\right)^{\frac{1}{\sigma_m}} \left(c_{jt} + \bar{c}_j\right)^{\frac{\sigma_m-1}{\sigma_m}}\right]^{\frac{\sigma_m}{\sigma_m-1}}, \text{ and} \\
c_{st} &= \left[\sum_{j\in I_s} \left(\omega_j^s\right)^{\frac{1}{\sigma_s}} \left(c_{jt} + \bar{c}_j\right)^{\frac{\sigma_s-1}{\sigma_s}}\right]^{\frac{\sigma_s}{\sigma_s-1}}
\end{aligned}$$

These equations define a set of nested preferences. The top nest is exactly as in Herrendorf, Rogerson, and Valentinyi (2013). Within each bottom nest, different goods are combined to form a bundle that is, in turn, an input to the top nest. Take, for example, the nest of manufactured products. I_m is the set of manufactured goods; σ_m parameterizes how easily different manufactured goods can be substituted for one another; and ω_j^m gives the importance of good j within the manufactured good bundle (where $\sum_{j\in I_m} \omega_j^m = 1$).

1. Download data from the NIPA tables (see Appendix A of the paper for the data sources). Define I_a, I_m, and I_s as follows:

   - I_a: Food and beverages purchased for off-premises consumption.
   - I_m:
     - Motor vehicles and parts
     - Furnishings and durable household equipment

- Recreational goods and vehicles
- Other durable goods
- Clothing and footwear
- Gasoline and other energy goods
- Other nondurable goods

• I_s:

- Housing and utilities
- Health care
- Transportation services
- Recreation services
- Food services and accommodations
- Financial services and insurance
- Other services
- Final consumption expenditures of nonprofit institutions serving households
- Government consumption expenditure

(a) Plot the quantity and price indices (from 1947 to the most recent period) for each of the industries in the services industry. Which service industry has had the largest price increase?

(b) Has consumption increased broadly across service industries, or is there substantial within-sector heterogeneity?

2. Write out the analogue of Equation (4) for industries within the "bottom nests." In other words, write out the expression for each service industry's expenditure as a share of overall expenditures within the service sector. And do the same for each manufacturing industry.

3. Estimate the equations that you have written in part 1 of this problem. (Hint: One of the authors, Akos Valentinyi, has posted the code related to this paper. See https://sites.google.com/site/valentinyiakos/Home/papers/st-preferences .) You will basically be running three sets of seemingly unrelated regressions: one for the top nest, one for industries within the service sector, and a third for industries within the manufacturing sector.

(a) Display your coefficient estimates. Within the service sector, which industries have the largest and smallest values of $\bar{c}_j$?

Problem 2

1. Read Sections 1-3 of http://scholar.harvard.edu/files/daniallashkari/files/clm_main.pdf (you can skip sections 3.4 and 3.6). What is the main criticism that this paper makes with regards to papers that use Stone-Geary preferences as a means of generating non-unitary income elasticities?

2. Download the dataset that is used in the paper (here:

   http://www.rug.nl/ggdc/productivity/10-sector/) . As best as you can, replicate their Table 1. It's ok if your results do not exactly align, but they should be in the right ballpark. Send me your code. (In general, I don't need to see your code, but it would be helpful for me in this instance.)

Problem 3

In this problem, we will explore changes wage income inequality, computing trends in "between-group" and "within-group" income inequality. For this problem, use your cleaned dataset for Problem 1 of Problem Set 1. Also, keep only data that begin in 1976.

Also, make the following adjustments and variable definitions. First, construct a categorical variable that takes five values, depending on the education of the worker: i) high school dropouts; ii) high school graduates; iii) some college; iv) college graduate; v) postgraduate education. Second, construct a categorical variable describing the years of potential labor market experience (equal to max[0,age−years of education−6]): i) 0-10 years of experience; ii) 11-20 years of experience; iii) 21-30 years of experience; and iv) 31+ years of experience. Third, construct an hourly wage equal to wage and salary income divided by (weeks worked last year × usual hours worked per week). In your sample, keep only individuals who have age between 16 and 65, have non-missing education and experience. Also, keep only workers who earn more than 90 percent of their total income from wage and salary income. Finally, remove observations for workers who have wages of less than \$1 or more than \$100 in 1979 dollars.

1. Compute and plot the trends, between 1976 and 2016, in the share of workers in each of your education groups, and in each of your experience groups. Do this separately for males and females.

2. Run the following regressions (separately for males and females)

$$\log w_{it} = \beta_0 + \beta_{\text{education}\times\text{experience}} + \beta_t + \epsilon_{it}$$

In this equation, the $\beta_{\text{education}\times\text{experience}}$ terms indicate that you should compute fixed effects for the $20 = 5 \times 4$ education $\times$ experience group variables. For each year, compute the variance of log wages and the variance of the residuals from your regression. The latter term is a measure of the "within-group" variance of wages. Call $\sigma^2(\log w_{it}) - \sigma^2(\epsilon_{it})$ the "between-group" variance. Plot the trends in "within-group," "between-group," and overall wage variance. Again, these plots should be separate for males and females.

3. Different education$\times$experience groups have different within-group variances. Which groups are these? Are these groups growing or shrinking as a share of the workforce over time?

For your information: As David Autor write in his lecture notes (here :

http://economics.mit.edu/files/7708), "There are two annual earnings series that are the source for much of what we know about the U.S. wage structure: the March Current Population Survey Annual Demographic File ('the March') and the May and (later) Monthly Outgoing Rotation Groups of the CPS. Although both are components of the Current Population Survey, they measure different earnings constructs. The March survey collects data on annual income whereas the May/MORG collects data on weekly or hourly income." In this problem, we were using the March CPS file. Comparing the two datasets, Lemieux (2006; "Increasing Residual Wage Inequality: Composition Effects, Noisy Data or Rising Demand for Skill?") demonstrates that, in the 1970s, within-group wage inequality increases much faster in (our) March CPS, and makes the case that wages in the Mar ORG file are better measured.