Research Professional Data Task

(Provided by Professor Matthew J. Notowidigdo)

This "Data Task" asks you to reproduce and extend Figure 4 and Table 2 in the Autor-Katz-Kearney (AKK) 2008 RESTAT paper that studies trends in income inequality in the United States. All of the data is provided in the "noto_data_task.csv" file (also provided as "noto_data_task.dta"). You can use whatever programming language you like (e.g., STATA, R, Python) to do the statistical analysis described below and create the figures and tables.

At the end of this task, you should submit your programming file, and you should create a single PDF file that explains your results and includes the figures and tables within the document. Submit your PDF here. We expect the task to take about 3-4 hours.

If anything is unclear about the assignment or the data that is provided, you should make whatever assumptions you need to be able to proceed and describe the assumptions in your write-up. You should begin by skimming the Autor-Katz-Kearney paper (paying closest attention to the discussion in Section III): https://economics.mit.edu/files/586.

Note that the data file includes a few additional years (up to 2008), so your figures and tables should go up through 2008 instead of 2005 as in AKK.

Here are the tasks:

1. Reproduce Figure 4 using all of the years in the data file. You are encouraged to use color in the figures and any other graphical formatting to make them as clear as possible.

2. Reproduce columns (1) through (5) of Table 2 using the same data file. Add a new row at the bottom of the table containing estimates, standard errors, and p-values of the elasticity of substitution (i.e., the reciprocal of the coefficient on the log relative supply variable).

3. Column (3) allows for a trend break in relative demand after 1992. Assume there exists a trend break at some point between 1964 and 2007; calculate and report the trend break year that maximizes the R^2 of the regression model. (That is, you should iterate through each year and assume that the trend break occurs in that year and calculate the R^2 of regression model that allows for trend break in that year, and find the year which maximizes the R^2 across all of the possible years.)