

1 Introduction

Thank you for applying to Professor Malmendier's research team.

This assignment is comprised of one task.

If you have questions that you think would be unanswerable by some other reasonable means, you may email me (Clint: cth@berkeley.edu)

2 Data Analysis Task

This task is a relatively brief data task. Those of you who apply for quantitative data based jobs and/or research assistant positions will encounter many of these in the future. Compared to the ones I've encountered in the past this one is much shorter and easier; however, I hope it will be a helpful learning experience for those of you trying your first data task and a helpful refresher to those who have completed one before.

Please try to complete as much of the assignment as you feel you are able.

I think it would probably be easiest to complete in Stata, through Berkeley you are provided with free access to Stata via Citrix. However, I created the solutions using Python so there is nothing that should prevent you from doing this in any reasonable application. I imagine that you will use Stata, Python, or R. If you want to use another application it is probably fine; however, please run it by me first (again, my email is cth@berkeley.edu).

Note: You are asked to provide your code.

You have been provided with data in the file `psid_task_data.txt`. This is a very well parsed set of data from the Panel Study of Income Dynamics. Normally, I would provide more detail about the data, but for this data task only the minimal information provided is needed.

The data set contains 5 columns:

- ID: This can be thought of as a household ID number
- change_in_fam_comp: This is a variable with numbers corresponding to different types of change in family composition
- year: This variable corresponds to the year
- full_income: This variable corresponds to the income of the household in a given year
- full_weights: This provides sample weights for the household in a given year

First, we want to impute birth years based on the data we have available. Please impute the birth year for all observations(don't overthink it, simple is fine). In your response please include the mean, median, and standard deviation for this variable.

Next, we would like you to provide a graph of average household income by year and output the underlying yearly data; however, there are a few things you need to do along the way.

1. We would like to use inflation adjusted income.
 - (a) To do this we would like to use the CPI for all urban consumers for all items (BLS data series CUUR0000SA0).
 - (b) This data is typically provided monthly, use the yearly average.
 - (c) Further, we would like our estimates to be in 2010 dollars; however, the series is based on 1982-1984. Please adjust the data series accordingly.
 - (d) Use this data to produce inflation adjusted income values.
2. We only want to consider households that have not had a recent change in family composition.
 - (a) Households that have not experienced a recent change have a "change_in_fam_comp" value of 0
 - (b) Exclude all observations where the household has experienced a recent change in family composition when calculating your average income estimates
3. We would like to weight the sample appropriately
 - (a) The full_weights variable provides sample weights
 - (b) Please use the weights appropriately when calculating average income

Bearing these three points in mind, please compute a graph of average income by year. More specifically, this will be a graph of average inflation adjusted income by year for households without a recent change in family composition where households have been appropriately waited.

Please include in your response a PDF with any relevant notes, your complete code, the raw CPI data you downloaded (so I can test your code), an image of the graph you produced, and the data set used to produce that graph (it should have 2 columns: year and average income).

Good Luck!