Empirical Task: Spatial Mobility in the NLSY79

The goal of this task is to evaluate your ability to efficiently extract, clearly present, and accurately interpret key information from a dataset. The task focuses on mobility patterns in the U.S. across the four main regions in the NLSY79. For each question, please use combinations of plots, tables, and linear regressions as you see fit. Please use the attached dataset to complete the following exercises in eight hours or less. You may use the programming language of your choice.

1 Summary Statistics

Report the count of moves across U.S. regions for each possible transition. Report the count of moves between urban and non-urban areas. Report mean wage income, mean employment, and mean educational attainment in each region and urban/non-urban bin. Comment on differences you might find.

2 Summarizing Data with Linear Regression

Consider the following target parameters:

(1) The average wage income difference in the 2004-2012 period between: (a) race-3 gender-2 urban workers in region 2; (b) race-1-or-2 gender-1 non-urban workers who reside anywhere outside region 2 (possibly inside the "NA" region).

(2) The average wage income earned by race-2 gender-1 non-urban workers in region 3 with an educational attainment in the 9-12 range.

(3) The average wage income difference between: (a) race-1 gender-1 workers in region 2 and born in 62; (b) urban workers in region 3, with an educational attainment in the 13-16 range, and in 2006.

For each of these target parameters:

(a) Propose and write a linear regression such that – using the *full sample* – a *single* coefficient captures that parameter. Indicate the specific regression coefficient that captures the target parameter.

(b) Report the estimated coefficient of interest and its standard error.

3 Event Studies

Using movers only, carry out a descriptive event study with wage income as the outcome. If you are not familiar with event studies, you may plot the mean wage income earned by movers in years leading to and following the move. In either case, please consider a relative time window between -2 and $+2$. In your plots, the x-axis should measure time relative to the move and the y-axis should display event-study estimates or mean wage income, depending on your choice above. Consider both moves between regions and moves between urban and non-urban areas.

4 Comparing Movers to Stayers

Compare movers' wage income with stayers' wage income both in the origin region and in the destination region. Do so while controlling for age, educational attainment, race, and gender. Carry out this comparison for different time periods relative to the move. Document and interpret differences in levels and differences in trends, if any.

Additional Information

About the data Columns: `i` is an individual identifier, `birth` is birth year, `gender` is a categorical variable for gender, `race` is a categorical variable for race, `region` is a categorical variable for geographical area, `urban` takes the value 1 if a worker lives in an urban area, `wage` is wage income, `year` is calendar year, and `educ` measures completed education. The dataset contains a lot of missing values for `region`, `urban`, and `wage`. Define employment as `wage > 0`. In sections 3 and 4, focus on individuals for which data before and after their move are nonmissing.

What to submit Please submit one report containing your results and analysis. I expect the report to be between 5 and 10 pages, including figures and tables. The text itself should not be longer than 3 pages. Please submit the code you used to generate results. Please provide comments within your code.