Econ 270 / GSB 603 **Department of Economics**

Fall 2025 **Stanford University**

Problem Set V

Due: Tuesday, November 11th, 2025, 6pm.

Be concise but clear as to what numbers you are reporting, and answer in full sentences. You should also hand in supporting code, but all answers should be in a PDF or Word document.

We use the LaLonde data, with 15,992 observations. We will use the data on earnings in 1978, earnings in 1975, earnings in 1974, education, indicators for Black and Hispanic, and age. (All earnings variables are coded as `reXX`, with `XX` being the year.)

1. Regress earnings in 1978 on earnings in 1975. What values for earnings in 1975 do you expect to have the highest leverage? What are the values for earnings in 1975 for the five highest leverage observations?
2. Do the regression also including earnings in 1974, education, indicators for Black and Hispanic, and age. Again, find the five observations with the highest leverage. Are the values of the leverage high enough to cause concern?
3. Calculate the residuals. Divide them by the standard deviation of the residuals. What are the highest five values in absolute value? Are these also observations with high leverage?
4. Consider the regression of earnings in 1978 on earnings in 1975. What is the out-of-sample root-mean-squared-error?
5. If you also include earnings in 1974, education, indicators for Black and Hispanic, and age in the regression, what is the out-of-sample root-mean-squared-error?