Introduction to Econometrics
Session 9 – Panel Data and Fixed Effects

December 2025

1 Problem

1. Load the `Males` dataset from the `plm` package. These data track 545 male full-time employees in the United States between 1980 and 1987.
2. For each individual, compute:
   - the difference between their average log wage (`wage`) during periods spent in jobs covered by collective bargaining (`union`) and those that are not;
   - the share of time spent in a job covered by collective bargaining;
   - the within-individual variance of the indicator representing whether they are in such a job.
3. What share of missing values do you observe for the individual mean wage difference? How can these missing values be explained?
4. Compute the sample average of the individual mean differences, weighting each individual proportionally to the variance of the indicator. What is the purpose of this weighting? For which population does this average provide an estimate?
5. Estimate, using the `lm` function, the regression of log wage on (i) the indicator for being in a job covered by collective bargaining and (ii) an indicator for each individual in the panel. How should the coefficient on `union` be interpreted? Compare this coefficient to the mean difference estimated in the previous question.
6. What is the coefficient in the regression of `union` on the share of time spent by each individual in a job covered by collective bargaining? Verify this by estimating the regression using the `lm` function.
7. Estimate the regression of log wage on `union` and the share of time each individual spends in a job covered by collective bargaining. Compare the coefficient on `union` to the previous results.

8. Use the `felm` function from the `lfe` package to estimate the regression of log wage on `union` with fixed effects at the individual level. Compare the coefficient on `union` to the previous results.

9. Which variance–covariance matrix should be used for this regression? Using the appropriate variance–covariance matrix, obtained directly with the `lfe` function, report the confidence interval for the coefficient.

10. Use the `felm` function to estimate the regression of log wage on education, measured as time spent in the schooling system (`school`), with individual fixed effects. What issue arises here?

11. Use the `felm` function to estimate the regression of log wage on experience (`exper`) and calendar year (`year`). What issue arises here?

12. Use the `felm` function to estimate the regression of log wage on `union` with fixed effects defined at the individual (`nr`) $\times$ industry (`industry`) level. How should the coefficient on `union` be interpreted?

13. Which variance–covariance matrix should be used for the previous estimation? Why?