Applied Data Analysis School
November 2021

Linear Panel Data Models
Exercise

The aim of this exercise is to estimate a mincerian wage equation using individual panel data from the National Longitudinal Survey (NLS) of young women aged 14 to 26 years in 1968 observed over several years. Consider the following model:

$$\begin{aligned} \ln\_wage_{it} &= \beta_1 collgrad_i + \beta_2 age_{it} + \beta_3 agesq_{it} + \beta_4 tenure_{it} + \beta_5 tensq_{it} + \\ &+ \beta_6 union_{it} + \beta_7 south_{it} + + \beta_8 not\_smsa_{it} + \beta_9 c\_city_{it} + \alpha_i + u_{it}, \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, ..., N; t = 1, ..., T \end{aligned}$$

Variables description:
ln_wage: log(wage);
collgrad: =1 if college graduate;
age: age in current year;
agesq: age squared;
tenure: job tenure (in years);
tensq: tenure squared;
union: =1 if belongs to a union;
south (residence): =1 if in the South area;
smsa: =1 if not in a standard metropolitan statistical area;
c_city: =1 if in a central city.
The data are available in the file *nlswork.dta*.

Tasks:
1. Estimate the model by pooled OLS (POLS).
2. Estimate the model by random effects (RE). Test for the presence of an unobserved effect.
3. Estimate the model by fixed effects (FE).
3.1 Test for the presence of fixed effects.
3.2 Test for the joint significance of *age* and *agesq*.
4. Compare the RE and FE estimators using the Hausman test.
5. Estimate the model using the between estimator (BE).
6. Estimate the model using the first differences estimator (FD).