Econometrics II: Assignment 2, Panel data models

The dataset NLSY2000RC_V2.dta(\.csv) is a paneldataset on male workers, working at least 30 hours a week. It has data for the years 1980-1994, 1996, 1998 and 2000. It contains the following variables:

ID	:	personal identifier
TIME	:	year $-$ 1980
EARNINGS	:	hourly wage
AGE	:	age
AGESQ	:	age squared
S	:	years of schooling
ETHBLACK	:	black ethnicity (dummy variable)
URBAN	:	living in urban area
REGNE	:	region north-east
REGNC	:	region north-central
REGW	:	region west
REGS	:	region south
ASVABC	:	index test score

The variable ASVABC is a composite test score for skills like arithmetic reasoning, word knowledge, paragraph comprehension. A higher value of ASVABC goes together with a higher test score. It is generally considered to be a good indicator for ability. It is measured only once, upon entering the panel, and therefore it is constant over time.

A researcher is interested in estimating the effect of schooling on the hourly wage. A basic equation is

$$\ln EARNINGS_{it} = \alpha_0 + \alpha_1 S_{it} + \alpha_2 AGE_{it} + \alpha_3 AGESQ_{it} + X'_{it}\gamma + U_{it}$$

where X includes remaining variables.

1. First use pooled OLS to check the impact of including and excluding ASVABC on the estimate of α_1. Present and explain the result.

The researcher is asked to analyze whether returns to schooling, as measured by the parameter α_1, is different for workers with a black ethnicity. A difference

in returns to schooling by ethnicity is sometimes interpreted as an indication of discrimination on the labour market. Two basic ways to model heterogeneity in the schooling parameter by ethnicity are (1) including a cross effect of schooling and ethnicity and (2) estimating separate equations by ethnicity.

2. Perform a pooled OLS analysis to obtain insight in the heterogeneity of returns to schooling by ethnicity. Present the results and comment on the outcomes: what are the conclusions based on this?

So far, the panel nature of the data has hardly been exploited. Random effects estimation can improve efficiency of the estimates compared to pooled OLS.

3. Perform the analysis for heterogenous schooling effects using the random effects model. Present the results and compare the outcomes with the pooled OLS results obtained before. Interpret the outcomes.

Alternatively, panel data can be exploited to perform fixed effects (or: within group) estimation.

4. A priori, would you plead for using fixed effects estimation or random effects estimation? Explain your answer.

5. Apply the fixed effects estimator to analyze the heterogenous schooling effect. Interpret the outcomes.

6. Fixed effects estimation may not be as efficient as random effects estimation, but is robust to correlation between regressors and the random effect. Can we perform a Hausman test in this context? Perform the test you propose.

7. Perform Mundlak estimation of the model. Present the results of estimation and test for the joint sigificance of the within-group means.

8. What are your overall conclusions from the analysis of heterogeneity in returns to schooling by ethnicity?

9. To gain insight in the impact of nonresponse and attrition, the researcher applies a variant of the Verbeek and Nijman-test (see lecture slides). He defines the dummy variable d_i which is 1 if the individual is in the panel for more than 5 waves, and is zero otherwise. Apply the Verbeek and Nijman test with this definition of d_i (otherwise equal to the definition at the lecture slides). Draw conclusions and address practical problems you possibly met in implementing the test.