Problem Set 05
Panel Data Models

TAs: Aida Hatami, Sam Fathinejad, Fatemeh Salehi, Reza Sahour

April 27, 2025

1. Labor Supply Panel Data Exercise

Use a 50% random subsample of the wage–hours data attached to your problem set (MOM data[1], variables are respectively `lnhr`, `lnwg`, `kids`, `ageh`, `agesq`, `disab`, `id`, `year`) and estimate β in the model

$$\ln hr_{it} = a_i + \beta \ln wg_{it} + E_{it}.$$

a) Can β be directly interpreted as a labor supply elasticity? Explain.

b) For the following estimators: (1) pooled OLS, (2) between, (3) within, (4) first differences, (5) random effects GLS, (6) random effects MLE, report (i) β (the estimated coefficient on $\ln wg$), (ii) the default standard error, and (iii) the panel-bootstrap standard error with 200 replications.

c) Are the estimates of β similar across estimators?

d) Is there a systematic difference between the default standard errors and the panel-robust standard errors?

e) Will the pooled OLS estimator in part (b) be consistent for β in a fixed-effects model? In a random-effects model?

f) Given the preceding evidence, do you believe that the labor supply curve is upward sloping? Explain.

2. WAGEPAN Data Analysis

Use the data in `WAGEPAN.RAW` to answer this question.

a) Using `lwage` as the dependent variable, estimate a model that contains an intercept and the year dummies `d81` through `d87`. Use pooled OLS, random effects (RE), fixed effects (FE), and first differences (FD) (in the latter case difference the year dummies and `lwage`, and omit the overall constant). What do you conclude about the coefficients on the dummy variables?

b) Add the time-constant variables `educ`, `black`, and `hisp` to the model, and estimate by pooled OLS and RE. How do the coefficients compare? What happens if you estimate by FE?

[1]The wage–hours panel dataset provided with this assignment

c) Are the pooled OLS and RE standard errors from part (b) the same? Which are probably more reliable?

d) Obtain fully robust standard errors for pooled OLS. Do you prefer these or the usual RE standard errors?

e) Obtain fully robust standard errors for RE. How do these compare with the fully robust OLS standard errors, and why?

3. LOWBIRTH Data Analysis

Use the data in `LOWBIRTH.RAW` for this question.

a) For 1987 and 1990, consider the state-level equation

$$lowbrth_{it} = \theta_1 + \theta_2\, d90_t + \beta_1\, afdcpcr_{it} + \beta_2\, \log\big(phypc_{it}\big) + \beta_3\, \log\big(bedspc_{it}\big) + \beta_4\, \log\big(pcinc_{it}\big) + \beta_5\, \log\big(popul_{it}\big) + c_i + u_{it},$$

where the dependent variable is the percentage of births classified as low birth weight and the key explanatory variable is `afdcpcr`, the percentage of the population in AFDC (Aid to Families with Dependent Children). The controls are physicians per capita, hospital beds per capita, per capita income, and population. Interpreting the equation causally, what sign should each β_j have? (Note: AFDC participation makes poor women eligible for nutritional programs and prenatal care.)

b) Estimate the preceding equation by pooled OLS and discuss the results. Report both the usual and serial-correlation–robust standard errors.

c) Difference the equation to eliminate the state fixed effects c_i and reestimate. Interpret the estimate of β_1 and compare it to that in part (b). What do you make of $\hat{\beta}_2$?

d) Add $\texttt{afdcpcr}^2$ to the model and estimate by FD. Are the estimates on `afdcpcr` and $\texttt{afdcpcr}^2$ sensible? What is the estimated turning point of the quadratic?