

## Aggregate Data Analysis

### Homework 1

Prf. José Fajardo (FGV/EBAPE)

1. Using two datasets *time\_var.dta* and *time\_invar.dta*, consider the following wage equation:

$$\ln \text{wage}_{it} = \alpha_0 + \alpha_1 \text{ability}_i + \alpha_2 \text{medu}_i + \alpha_3 \text{fedu}_i + \alpha_4 \text{d}_i + \alpha_5 \text{siblings}_i + \beta_1 \text{e}_{it} + \beta_2 \text{pexp}_{it} + \varepsilon_{it}$$

Notice that all the  $\alpha$  coefficients are associated with time-invariant cross section data, while  $\beta$  are with time-variant panel data series.

- a) Formulate, estimate, and compare the pooled or population-averaged based on OLS and OLS with panel-robust standard errors, respectively. In addition to pooled model, three different variable transformations should be considered and compared: (1) first-difference, (2) between (or group means), and (3) within (or deviations from group means). Note: not all coefficients can be estimated for all models. Why?
- b) Formulate, estimate, and compare the fixed-effects and random-effects panel data models based on OLS and OLS with panel-robust standard errors, respectively. Setup and perform hypothesis testings to choose a proper panel data model: (1) pool or not to pool? (2) fixed-effects or random-effects?

If you are interested in the original paper below, read [this](#), but we are not attempting to replicate their results (see also Joshua C. C. Chan, "Replication of the Results in 'Learning about Heterogeneity in Returns to Schooling'", Journal of Applied Econometrics, Vol. 20. No. 3, 2005, pp. 439-443.)

[Koop, G. and J. Tobias, "Learning About Heterogeneity in Returns to Schooling." Journal of Applied Econometrics, 19, 2004, 827-849]

2. The data in the file *productivity.txt* are a panel on the following variables for the lower 48 states, 17 years,

STATE = state name  
YR = year, 1970,...,1986  
P\_CAP = public capital  
HWY = highway capital  
WATER = water utility capital  
UTIL = utility capital  
PC = private capital  
GSP = gross state product  
EMP = employment  
UNEMP = unemployment rate

The basic model of interest is

$$Y_{it} = \beta_1 X1_{it} + \beta_2 X2_{it} + \beta_3 X3_{it} + \beta_4 X4_{it} + \beta_5 X5_{it} + c_i + \varepsilon_{it}$$

Where Y is logGSP, X1 is logPC, X2 is logHWY, X3 is logWATER, X4 is logUTIL and X5 is logEMP.

This is a Cobb-Douglas production function.

- a) Fit the “pooled” model and report your results
- b) Fit a random effects model and a fixed effects model. Use your model results to decide which is the preferable model. If you find that neither panel data model is preferred to the pooled model, show how you reached that conclusion

[Munnell, A. "Why Has Productivity Declined? Productivity and Public Investment." *New England Economic Review*, 1990, 3-22.]

**PS: Deadline February 20<sup>th</sup>, 2018**



## Aggregate Data Analysis

### Homework 1

Prf. José Fajardo (FGV/EBAPE)

1. Using two datasets *time\_var.dta* and *time\_invar.dta*, consider the following wage equation:

$$\ln \text{wage}_{it} = \alpha_0 + \alpha_1 \text{ability}_i + \alpha_2 \text{medu}_i + \alpha_3 \text{fedu}_i + \alpha_4 \text{d}_i + \alpha_5 \text{siblings}_i + \beta_1 \text{e}_{it} + \beta_2 \text{pexp}_{it} + \varepsilon_{it}$$

Notice that all the  $\alpha$  coefficients are associated with time-invariant cross section data, while  $\beta$  are with time-variant panel data series.

- a) Formulate, estimate, and compare the pooled or population-averaged based on OLS and OLS with panel-robust standard errors, respectively. In addition to pooled model, three different variable transformations should be considered and compared: (1) first-difference, (2) between (or group means), and (3) within (or deviations from group means). Note: not all coefficients can be estimated for all models. Why?
- b) Formulate, estimate, and compare the fixed-effects and random-effects panel data models based on OLS and OLS with panel-robust standard errors, respectively. Setup and perform hypothesis testings to choose a proper panel data model: (1) pool or not to pool? (2) fixed-effects or random-effects?

If you are interested in the original paper below, read [this](#), but we are not attempting to replicate their results (see also Joshua C. C. Chan, "Replication of the Results in 'Learning about Heterogeneity in Returns to Schooling'", Journal of Applied Econometrics, Vol. 20. No. 3, 2005, pp. 439-443.)

[Koop, G. and J. Tobias, "Learning About Heterogeneity in Returns to Schooling." Journal of Applied Econometrics, 19, 2004, 827-849]

2. The data in the file *productivity.txt* are a panel on the following variables for the lower 48 states, 17 years,

STATE = state name  
YR = year, 1970,...,1986  
P\_CAP = public capital  
HWY = highway capital  
WATER = water utility capital  
UTIL = utility capital  
PC = private capital  
GSP = gross state product  
EMP = employment  
UNEMP = unemployment rate

The basic model of interest is

$$Y_{it} = \beta_1 X1_{it} + \beta_2 X2_{it} + \beta_3 X3_{it} + \beta_4 X4_{it} + \beta_5 X5_{it} + c_i + \varepsilon_{it}$$

Where Y is logGSP, X1 is logPC, X2 is logHWY, X3 is logWATER, X4 is logUTIL and X5 is logEMP.

This is a Cobb-Douglas production function.

- a) Fit the “pooled” model and report your results
- b) Fit a random effects model and a fixed effects model. Use your model results to decide which is the preferable model. If you find that neither panel data model is preferred to the pooled model, show how you reached that conclusion

[Munnell, A. "Why Has Productivity Declined? Productivity and Public Investment." *New England Economic Review*, 1990, 3-22.]

**PS: Deadline February 20<sup>th</sup>, 2018**

