Please hand in your solutions on Tuesday, May 19, before the tutorial.

Exercise 1: Panel data model

The between model is a special case of the pooled model where we average each component across all time periods, i.e. $\bar{q}_i = \frac{1}{T}\sum_{t=1}^{T} q_{it}$. Consider the following between model with a single explanatory variable $\bar{x}_i$:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i \tag{1}$$

α_i and $\bar{u}_i$ are unobserved. Since a constant is included, it is reasonable to assume that $\mathrm{E}(a_i) = 0$ and $\mathrm{E}(\bar{u}_i) = 0$. Further assume that $\bar{x}_i$ is orthogonal to $\bar{u}_i$. However, $Cov(x_{it}, a_i) = \sigma_{xa}$.

(a) The between estimator $\hat{\beta}_1^B$, is the OLS estimator of equation (2), such that we obtain the following probability limit:

$$\text{plim } \hat{\beta}_1^B = \beta_1 + \tfrac{\sigma_{xa}}{Var(\bar{x}_i)} \quad \text{if } N \to \infty$$

Show this!

(b) If x_{it} is uncorrelated over time ($t = 1, ...T$) and constant across individuals, i.e. $Var(x_{it}) = \sigma_x^2$, the probability limit of exercise (a) simplifies to:

$$\text{plim } \hat{\beta}_1^B = \beta_1 + T\tfrac{\sigma_{xa}}{\sigma_x^2} \quad \text{if } N \to \infty.$$

Show this!

(c) Given that the previous assumptions hold, describe verbally how the inclusion of additional time periods affects the consistency of the between estimator.
Is the assumption that x_{it} is uncorrelated over time reasonable?

Exercise 2: FE vs LDV

(a) You want to analyze the impact of training measures on wages. For this means you have collected a panel data set of working individuals over some years. You expect the impact of wages to be persistent over time and you are worried about unobserved heterogeneity that might impact your outcome measure as well as the assistance of training measures. To overcome both concerns your want to work with a model that includes a lagged dependent variable of your outcome variable as well as individual fixed effects. Your specification reads:

$$y_{it} = y_{i,t-1}\beta_1 + X_{it}\beta_2 + \delta D_{it} + \alpha_i + \tau_t + u_{it} \qquad \text{for } t = 1, ..., 5.$$

Where y_{it} denotes the outcome measure, i.e. the wage in year t. D_{it} is a binary variable indicating whether individual i participated in a training measure in period t. By the inclusion of a set of individual fixed effects (α_i), the lagged dependent outcome variable ($y_{i,t-1}$), year dummies (τ_t) and further individual specific explanatory variables (X_{it}) you want to estimate δ by simple OLS. Can δ be consistently estimated? Explain.

(b) Have look at Appendix 5.4 of *Mostly Harmless Econometrics* and explain how Angrist and Pischke deal with the previous problem and what they suggest an applied econometrician to do in a similar situation.

Exercise 3: First Differences

You want to investigate the impact of enterprise zones (EZ) on unemployment claims (UNEM). To do so, you draw on the instructional dataset EZUNEM.dta provided by Wooldridge. You can directly open the data-set in Stata using the following command „*bcuse ezunem*“. In case of an error message you want to instal the ado-file *bcuse* first („*ssc install bcuse*“) . A description of all included variables can be found here: `http://fmwww.bc.edu/ec-p/data/wooldridge/ezunem.des`. Starting point of your analysis is the following model:

$$log(uclms_{i,t}) = \beta_1 ez_{i,t} + \alpha_i + \tau_t + \epsilon_{i,t} \tag{2}$$

with $uclms_{i,t}$ denoting the number of unemployment claims filed during year t in city i. α_i and τ_t denote a set of city and year dummies.

(a) To eliminate time-constant characteristics of a city, write down the model in First Differences. In a second step, transform the respective variables in Stata and estimate your differenced model. How do you interpret the coefficient of interest ($\hat{\beta}_1$)? Is there any need to adjust the standard errors?

(b) You are concerned that cities follow different time trends. Since this might bias your result, you consider the following model:

$$log(uclms_{i,t}) = \beta_1 ez_{i,t} + \alpha_i + \alpha_i * year_t + \epsilon_i \tag{3}$$

With the interaction $\alpha_i * year_t$ you account for the possibility that each city follows its own linear time trend. In a fist step, you take First Differences of equation (3). Write down the respective model in First Differences and estimate this model with Stata in subsequent step.

(c) Include to your First Differences model that you derived in (b) a full set of year dummies? How does β_1 compare to (a) and (b)?