Exercise 1: Panel data models

Consider the following panel data model with one regressor w_{it} and two time periods.
$y_{it} = \beta_1 + w_{it}\beta_2 + \alpha_i + \epsilon_{it}, \qquad i = 1, ..., n, \qquad t = 1, 2.$

We assume that:

$$\begin{pmatrix} \alpha_i \\ w_{i1} \\ w_{i2} \\ \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim iid\mathcal{N} \left[\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \sigma_\alpha\sigma_w/2 & \sigma_\alpha\sigma_w/2 & \sigma_\alpha\sigma_\epsilon/3 & \sigma_\alpha\sigma_\epsilon/3 \\ \sigma_\alpha\sigma_w/2 & \sigma_w^2 & \sigma_w^2/2 & 0 & 0 \\ \sigma_\alpha\sigma_w/2 & \sigma_w^2/2 & \sigma_w^2 & 0 & 0 \\ \sigma_\alpha\sigma_\epsilon/3 & 0 & 0 & \sigma_\epsilon^2 & 0 \\ \sigma_\alpha\sigma_\epsilon/3 & 0 & 0 & 0 & \sigma_\epsilon^2 \end{pmatrix} \right]$$

(a) Does this model satisfy the random effects assumptions? Is the random effects OLS estimator that regresses y_{it} on a constant and w_{it} consistent for β_1 and β_2?

(b) We now want to estimate the model using the fixed effects assumption on α_i. Can β_1 be consistently estimated? Can β_2 be consistently estimated?

Exercise 2: A further question on panel data models

Consider the static panel data model

$$y_{it} = \beta x_{it} + \epsilon_{it},$$

where $i = 1, ..., n$ denotes individuals, and $t = 1, 2$ denotes time. The single regressor x_{it} and the error term ϵ_{it} are distributed as follows

$$\begin{pmatrix} x_{i1} \\ x_{i2} \\ \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \sim iid\mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 2 & 2 & 2 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{pmatrix} \right]$$

For any real number q define the differences $\Delta^{(q)}y_i = y_{i1} - qy_{i2}$, $\Delta^{(q)}x_i = x_{i1} - qx_{i2}$, $\Delta^{(q)}\epsilon_i = \epsilon_{i1} - q\epsilon_{i2}$. For $q = 1$ these are standard first differences. The OLS estimator of the regression of $\Delta(q)y_i$ on $\Delta(q)y_i$ is given by

$$\hat{\beta}^{(q)} = \frac{\sum_{i=1}^n \Delta^{(q)}y_i\Delta^{(q)}x_i}{\sum_{i=1}^n (\Delta^{(q)}x_i)^2}$$

(a) Is the OLS estimator of the regression of y_{it} on x_{it} consistent for β? Explain.

(b) Show that there exists a constant c such that

$$x_{i1} = c\alpha_i + \tilde{x}_{i1}, \qquad x_{i2} = \alpha_i + \tilde{x}_{i2}, \qquad \epsilon_{i1} = \alpha_i + u_{i1}, \qquad \epsilon_{i2} = \alpha_i + u_{i2}$$

where α_i, $\tilde{x}_{i1}$, $\tilde{x}_{i2}$, u_{i1} and u_{i2} are mutually independent standard normal random variables. What is the value of c? (Hint: you can start with the representation of x_{i1}, x_{i2} , ϵ_{i1} and ϵ_{i2} given here. Then show that x_{i1}, x_{i2}, ϵ_{i1} and ϵ_{i2} are jointly normal with mean zero, and that there exists c that delivers the 4×4 variance-covariance matrix above).

(c) For $q = 1$, explain why $\hat{\beta}^{(1)}$ is consistent and asymptotically normal, and derive the asymptotic variance of $\hat{\beta}^{(1)}$.

Exercise 3: *Get familiar with Stata*

The purpose of this exercise consists in familiarizing yourself with Stata.

Download the CPS data on wages from 1978 and 1985 using the following link `https://www.stata.com/data/jwooldridge/eacsap/cps78_85.dta`.[1] Using the „y85“-dummy restrict the data set to the year 1985. Thus, you end up with one cross section including 533 randomly selected employed workers.

Throughout the exercise use robust standard errors.

(a) Consider the model:

$$lwage_i = \alpha + \beta_1 female_i + \beta_2 union_i + \beta_3 nonwhite_i + \beta_4 educ_i + \beta_5 exper_i + \beta_6 exper_i^2 + u_i$$

Estimate the model by OLS. Where does the identifying variation come from? Could you account for time constant unobservables if an individual identifier was available?

How does the coefficient on union-membership interpret? Assume that you have a panel data set with individual identifiers available and that you want to estimate the impact of union-membership using an individual FE-approach. What kind of variation do you need to identify the impact of union-membership?

(b) What is the marginal effect of experience? At which level of experience is *lwage* maximized in this model? Plot the experience-earnings profile.

(c) In this data set the variable *exper* is potential experience, i.e. it is defined as *age* minus *educ*. Can we include *age* as an additional regressor in the above model? Explain and show what Stata does.

[1] Population survey conducted by the Department of Commerce in the US. The teaching data set provided by Jeffrey Wooldridge.

(d) If individuals with more schooling receive more on-the-job training, then we would suspect that the experience-earnings profiles of the more educated should be steeper than those of the less educated. How would you test this?