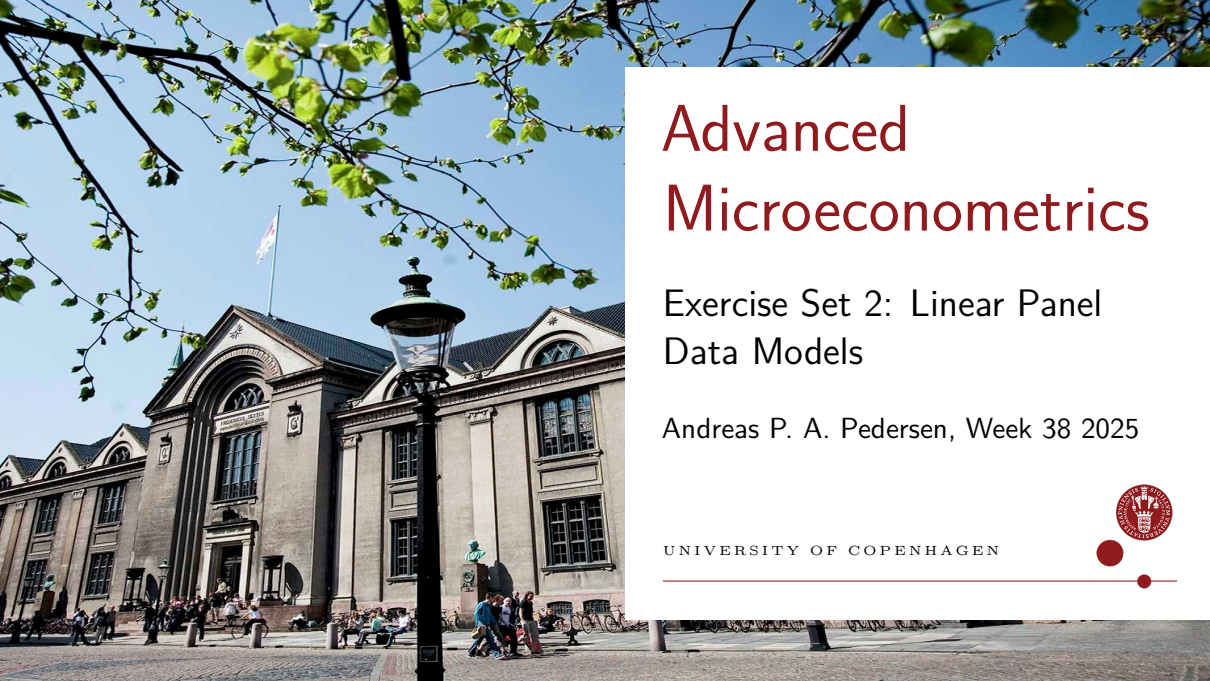

Advanced Microeconometrics

Exercise Set 2: Linear Panel Data Models

Andreas P. A. Pedersen, Week 38 2025

UNIVERSITY OF COPENHAGEN

Today's plan

- Linear Panel Data Models
- Perm(utation) Function
- Tests for Strict Exogeneity and Serial Correlation
- A few notes on good practice for assignment
- Your time to shine!

General Panel Data Model

- General model:

$$y_{it} = c_{it} + \mathbf{x}_{it}\beta + u_{it} \tag{1}$$

- How can we consistently estimate β_{it}? Which assumptions do we require for consistency?
  - Homoskedasticity
  - Exogeneity assumption(s)
- For the exogeneity assumption we need to make assumptions about c_{it} and especially $E(c_{it}\mathbf{x}_{it})$ as well as $E(u_{it}\mathbf{x}_{it})$

Linear Panel Data Model

- Linear model:

$$y_{it} = c_i + \mathbf{x}_{it}\beta + u_{it} \tag{2}$$

  where $\beta_{it} = \beta$ and $c_{it} = c_i \ \forall$ i,t and c_i unobserved
- Exogeneity: even if we believe an exogeneity assumption $E(u_{it}\mathbf{x}_{it}) = 0$ concerning u_{it} is plausible. If $E(c_{it}\mathbf{x}_{it}) \neq 0$, then $\hat{\beta}$ is inconsistent under POLS
- Solution: allow for arbitrary correlation between c_i and $\mathbf{x}_{it}$ by demeaning or first differencing individual level data and applying OLS to the transformed data. Then we only need the strict exogeneity assumption on u_{it}

FE & FD Assumptions

- Fixed Effects Assumptions:
  1. (FE): $E(\mathbf{x}_{it} u_{is}) = 0 \ \forall \ i \in N$ and $s, t \in T$
  2. (FE): rank$(E(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})) = \mathrm{K} \ \forall \ i \in N$
  3. (FE): $E(\mathbf{u}_i \mathbf{u}_i') = \sigma_u^2 U_T \ \forall \ i \in N$
- First Difference Assumptions:
  1. (FD): $E(\mathbf{x}_{it} u_{is}) = 0 \ \forall \ \in N$ and $s, t \in T$
  2. (FD): rank$(\mathrm{E}(\Delta\mathbf{X}'\Delta\mathbf{X})) = \mathrm{K} \ \forall \ i \in N$
  3. (FD): $\mathrm{E}(\mathbf{e}_i \mathbf{e}_i') = \sigma_e^2 I_{T-1} \ \forall \ \in N$ where $\mathbf{e}_i = \Delta\mathbf{u}_i$
- Homoskedasticity = homogeneous distribution of error terms

Perm(utation) Function

- To transform the data we use the demeaning matrix Q and the differencing matrix D:

$$\mathbf{Q} = I_T - \begin{bmatrix} \frac{1}{T} & \cdots & \frac{1}{T} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} & \cdots & \frac{1}{T} \end{bmatrix}_{T \times T} \tag{3}$$

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}_{T-1 \times T} \tag{4}$$

- The perm function applies these matrices onto the data sets, returning the transformed dataset. We then perform regressions on the transformed data.

Test for strict exogeneity (FE.1/FD.1)

- Under the strict exogeneity assumption a lead of an explanatory variable $\mathbf{x}_{it+1}$ should be insignificant, conditional on that variable $\mathbf{x}_{it}$ being included in the estimating equation (why?)
- A test of strict exogeneity (Wooldridge p.325) when T $>$ 2: A) Run an FE estimation:

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{w}_{it+1}\delta + c_{it} + u_{it} \tag{5}$$

B) Test the null hypothesis that $\delta = 0$ (strict exogeneity) vs $\delta \neq 0$ using either a t-test or a Wald test

Test for serial correlation (FD.3)

- Under FD.3, the first difference errors $e_{it} = \Delta u_{it}$ are homoskedastic and therefore serially uncorrelated $E(e_{it}e_{it-1})$=0
  - If the idiosyncratic errors u_{it} follow a random walk $u_{it} = u_{it-1} + e_{it}$ then e_{it} will be serially uncorrelated
  - If the idiosyncratic errors u_{it} are uncorrelated then Corr$(e_{it}e_{it-1})$ = -0.5
- Assuming strict exogeneity (FE.1/FD.1) holds, a test of serial correlation in e_{it} can be performed as a test of whether $\gamma = 0$ for the regression:

$$\hat{e}_{it} = \gamma \hat{e}_{it-1} + v_{it} \tag{6}$$

  With (usable) periods T-2 (Wooldridge p.320)

- **Caution:** Failure to reject serial correlation in $\hat{e}_{it}$ does not exclude other causes of FD.3 failing

Assignment

- State the purpose and conclusions of your tests clearly!
- (Minimal) Testing Checklist:
  - What are the null and alternative hypotheses?
  - Which significance level are testing at?
  - Under what conditions is your test valid?
  - What can (and can't) you conclude from your test?
- Explain relevant notation
- Tables and graphs should be as self-explanatory as possible
- Complete referencing
- Implicate economic intuition in econometric analysis

Chat Bots

- University policy has changed to full integration of Chat bots in all courses.
- In my experience, they will be mostly correct most of the time, although will leave out essential details or give incorrect reasoning. -> Reflection of contaminated dataset, that is, even professionals make mistakes on the material of this course regularly.

Your time to shine!

- Read the entire problem set - highly applicable
- Make sure to correctly fill out the toolbox in w2_LinearModels_ante.py - otherwise you will get wrong estimates in the ipynb file
- Solve the rest of the problem set and use functions from the toolbox where necessary
- Many of the questions require you to look-up functions (.py file or documentation), theory (book, pdf or lecture slides)
- Paying attention to the transformations gives you most of the answers
- Use these slides for reference