Econometrics II

Lecture 4
Panel Data Models II

Marie Paul

University of Duisburg-Essen
Ruhr Graduate School in Economics

Summer Semester 2020

Outline of lecture

4. Panel Data Models II
   - The Correlated Random Effects Approach
   - Dynamic Panel Data Models
   - Mixed Linear Models

Self-study plan for Lecture IV

- This lecture discusses some more advanced panel data models.
- For our next virtual meeting, please work through all slides of Lecture IV.
- Correlated Random Effects Approach: Wooldridge (2019): p.474-476 (Chapter 14.3).
- Dynamic Models: Cameron and Trivedi: Chapter 22.5. Dynamic Models (from beginning until the end of Arellano-Bond) and Angrist and Pischke: Chapter 5.3.
- Mixed Linear Models: Cameron and Trivedi: Chapter 22.8
- You do not necessarily need to read the paper from the application (Paul, 2016).
- As usual, please prepare the "your own research" questions and collect questions and topics to discuss them in our virtual meeting.

Outline

4. Panel Data Models II
   - The Correlated Random Effects Approach
   - Dynamic Panel Data Models
   - Mixed Linear Models

Correlated Random Effects

The following slides are based on Wooldridge 2019:
As an alternative to fixed effects the approach **models the correlation** between α_i and $\mathbf{x}_{it}$. For simplicity take an individual effects model with only one regressor:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \tag{4.1}$$

Assume a simple **linear relationship**:

$$\alpha_i = \delta + \gamma \bar{x}_i + r_i \tag{4.2}$$

where we assume r_i uncorrelated with each x_{it}, Substituting gives:

$$y_{it} = \delta + \beta_{CRE} x_{it} + \gamma \bar{x}_i + r_i + \varepsilon_{it} \tag{4.3}$$

Note that ε_{it} is uncorrelated to $\bar{x}_i$ and for r_i the random effect assumption holds.
$\bar{x}_i$ controls for the correlation between α_i and the xses.
So we have a **random effects equation** with the addition of $\bar{x}_i$.

Correlated Random Effects

- It turns out that $\beta_{CRE} = \beta_{FE}$. Adding the time average $\bar{x}_i$ and using random effects is the same as substracting the times average and using POLS.
- This is a useful **interpretation of FE**: it **controls for the average level** when measuring the partial effect.
- This also gives an intuition why FE often suffers from **large standard errors**: if x_{it} and $\bar{x}_i$ are highly correlated the partial effect is difficult to estimate (multicollinearity).
- If γ is significant this **rejects the RE model** in favour of FE.

Adding Time Constant Regressors

- In contrast to a FE model, in a CRE model **time constant regressors** may be added:

$$y_{it} = \delta + \beta_{CRE} x_{it} + \gamma_1 \bar{x}_i + \gamma_2 z_i + r_i + \varepsilon_{it} \tag{4.4}$$

- β_{CRE} is still the FE estimate from the equation without z_i. In addition we obtain γ_2 but this does not necessarily represent a **causal** effect of z on y.
- The results also hold for a model with **multiple time-varying and time-constant regressors**.
- While the CRE model plays a minor role in linear panel analysis, it is important when it comes to **non-linear models**.

Outline

4. Panel Data Models II
   - The Correlated Random Effects Approach
   - Dynamic Panel Data Models
   - Mixed Linear Models

Dynamic Models: Definition

Consider again the usual individual-specific effects panel data model, but where a one period **lagged dependent variable** is entering as an additional regressor

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + \alpha_i + \varepsilon_{it}. \tag{4.5}$$

Dynamic models of the type in (4.5) are motivated whenever we expect the dependent variable to exercise some kind of **persistence** over time. For example:

- **Unemployment:** Individuals that become unemployed might, due to their unemployment status, be less attractive to employers (scarring effects).
- **Individual earnings:** Wage bargaining usually takes place using the employees' last years earnings as basis for negotiation.
- **Health status:** In follow-up surveys, individuals may anchor their subjective health status based on last periods answer (e.g., better/worse).

Such **state dependency** across time is likely to be prevalent in many practical applications and will cause severe problems for consistency as we will see.

Dynamic models: State Dependence and Unobserved Heterogeneity

We first concentrate on the case with $\boldsymbol{\beta} = \mathbf{0}$ so that (4.5) becomes

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, \tag{4.6}$$

with the additional restrictions that $|\gamma| < 1$ and ε_{it} are serially uncorrelated.

Both γ and α_i introduce correlation over time in y_{it}, but in **distinct** ways. To see this, note that under the assumptions on (4.6)

$$\begin{aligned}
\rho_{\Delta_{y_{it}}} = \text{Corr}[y_{it}, y_{i,t-1}] =& \text{Corr}[\gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}, y_{i,t-1}] \\
=& \gamma + \text{Corr}[\alpha_i, y_{i,t-1}] \\
=& \gamma + \frac{(1-\gamma)}{1 + (1-\gamma)\sigma_\varepsilon^2/(1+\gamma)\sigma_\alpha^2}.
\end{aligned} \tag{4.7}$$

(The latter holds in case of a random effects structure.)

Dynamic models: State Dependence and Unobserved Heterogeneity

Hence, equation (4.7) shows that correlation can arise from two sources:

1. **True state dependence:** $\alpha_i = 0$ or $\sigma_\alpha^2/\sigma_\varepsilon^2 \simeq 0 \rightarrow \rho_{\Delta_{y_{it}}} \simeq \gamma$. There is a causal mechanism that y_{it-1} determines y_{it}.
2. **Unobserved heterogeneity:** $\gamma = 0 \rightarrow \rho_{\Delta_{y_{it}}} = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$. Correlation of past and current outcomes due to unobserved individual effect. Correlation in error term would vanish if adding observables would make α_i vanish.

State Dependence and Unobserved Heterogeneity

Example 4.1 (Effect of public-sponsored training on earnings)

$$log(wage)_{it} = \gamma log(wage)_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} \tag{4.8}$$

Yearly data on employee's earnings. Here $\mathbf{x}_{it}$ includes dummies if a person i participates in public sponsored training in year t, in year $t-1$ and so on plus control variables.

α_{it} represents ability, motivation etc. and this influences wage and is potentially also correlated to training due to (self-)selection into training.

- **Unobserved heterogeneity**: e.g. low unobserved ability
- **State dependence**: e.g. once in unemployment hurdles are higher to obtain a job than to keep one
- **Duration dependence**: Additional lags may be relevant: the longer under low earnings, the more difficult to obtain a qualified job.

State Dependence and Unobserved Heterogeneity

Example 4.2 (Your own research)

Does state dependence play a role in your research?

Dynamic models: Inconsistency of Panel Estimators

All estimators from last lecture are **inconsistent** if the regressors include lagged dependent variables.

The problem is due to the fact that $y_{i,t-1}$ is correlated with the error term since

$$y_{i,t-1} = \gamma y_{i,t-2} + \mathbf{x}'_{i,t-1}\boldsymbol{\beta} + (\alpha_i + \varepsilon_{i,t-1}). \tag{4.9}$$

- **Pooled OLS:**

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + (\alpha_i + \varepsilon_{it}), \tag{4.10}$$

  where $\mathrm{E}[(\alpha_i + \varepsilon_{it})|y_{i,t-1}] \neq 0$ violates assumption POLS.1. *But note: OLS is consistent if errors are serially uncorrelated in a lagged dependent variables model without an individual effect (pooled model).*

- **Fixed Effects:**

$$(y_{it} - \bar{y}_i) = \gamma(y_{i,t-1} - \bar{y}_{i,t-1}) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i), \tag{4.11}$$

  where $\mathrm{E}[\varepsilon_{i,t-1}|y_{i,t-1}] \neq 0 \rightarrow \mathrm{E}[(\varepsilon_{it} - \bar{\varepsilon}_i)|(y_{i,t-1} - \bar{y}_{i,t-1})]$ violates assumption FE.1 (**strict exogeneity** assumption)

- **Random Effects:** Linear combination of FE and POLS $\rightarrow$ assumptions violated.

Dynamic Models: Consistent Estimators – IV

The **first-differenced** version of the dynamic model (4.5) is

$$(y_{it} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}). \tag{4.12}$$

Also the FD estimator is inconsistent since the correlation between $y_{i,t-1}$ and $\varepsilon_{i,t-1}$ in (4.9) implies that $(y_{i,t-1} - y_{i,t-2})$ is correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$.

However, a suitable **instrument** for the endogenous $(y_{i,t-1} - y_{i,t-2})$ is $y_{i,t-2}$ since:

1. $y_{i,t-2}$ is clearly correlated with $(y_{i,t-1} - y_{i,t-2})$ so it is **relevant**.
2. $y_{i,t-2} = \gamma y_{i,t-3} + \mathbf{x}'_{i,t-2}\boldsymbol{\beta} + (\alpha_i + \varepsilon_{i,t-2})$ is uncorrelated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$ so it is **valid**.

A critical assumption for this approach to work is that the errors ε_{it} cannot be **serially correlated**; otherwise the instrument will not be valid. Unrealistic in many settings! E.g. People's earnings are highly correlated from one year to the next due to wage agreements holding for two years for example.

Dynamic Models: Consistent Estimators – GMM

The idea of using a lagged regressor as an instrument for itself can be extended to using additional lags of the dependent variable as instruments. The higher the period the more instruments are available:

The **Arellano-Bond estimator** is defined as the resulting GMM IV estimator from the first differenced model (4.12) with moment condition $\mathrm{E}[y_{is}\Delta u_{it}]$ for $s \leqslant t-2$

$$\hat{\boldsymbol{\beta}}_{AB} = \left[\left(\sum_{i}^{N} \widetilde{\mathbf{X}}_i' \mathbf{Z}_i\right) \mathbf{W}_N \left(\sum_{i}^{N} \mathbf{Z}_i' \widetilde{\mathbf{X}}_i\right)\right]^{-1} \left(\sum_{i}^{N} \widetilde{\mathbf{X}}_i' \mathbf{Z}_i\right) \mathbf{W}_N \left(\sum_{i}^{N} \mathbf{Z}_i' \widetilde{\mathbf{y}}_i\right) \tag{4.13}$$

where $\widetilde{\mathbf{X}}_i$ is a $(T-2) \times (K+1)$ matrix with tth row $(\Delta y_{i,t-1}, \Delta \mathbf{x}_{it}')$ and $\mathbf{Z}_i$ is a $(T-2) \times r$ instrument matrix with components $\mathbf{z}_{it}' = [y_{i,t-2}, y_{i,t-3}, \ldots, y_{i1}, \Delta \mathbf{x}_{it}']$

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i3}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_{iT}' \end{bmatrix}. \tag{4.14}$$

Bracketing Property of FE and Lagged Dependent Variables

Angrist and Pischke (2009): If the **lagged dependent variables model** is the correct one and you use FE, estimation of positive treatment effects tends to be too large.

If **FE** is the correct model and you use lagged dependent variables your estimates tend to be too small.

Example 4.3 (Public sponsored training)

Earnings of participants may be low because of low qualification (time-constant, FE) or because they recently experienced a setback in their earnings (Ashenfelters dip). The latter is not time-constant and thus not a fixed-effect but may be controlled for (important if it has a lasting effect on wage) in a lagged dependent variables model.

Outline

4. Panel Data Models II
   - The Correlated Random Effects Approach
   - Dynamic Panel Data Models
   - Mixed Linear Models

Mixed Linear Models: Definition

In the last lecture we defined the general panel data model for cross section unit i in time period t as

$$y_{it} = \boldsymbol{\alpha}_{it} + \mathbf{x}'_{it}\boldsymbol{\beta}_{it} + \varepsilon_{it}, \tag{4.15}$$

We focused on the special case when the **intercept**, α, was individual-specific and the slope was constant, i.e., $\alpha_{it} = \alpha_i$ and $\beta_{it} = \beta$.

However, richer linear models additionally permit also the **slope parameters** β to vary across observations.

Such models are usually called **mixed linear models** but also **mixed effects**, **hierarchical**, **random coefficients**, or **variance components** models.

Importantly, these models usually assume (as in the random effects model) that there are **no fixed effects** present, i.e. $\mathrm{E}[\alpha_i|\mathbf{x}_{it}] = \mathrm{E}[\alpha_i] = 0$.

Mixed Linear Models: Specification

The mixed linear model specifies

$$y_{it} = \mathbf{z}_{it}'\boldsymbol{\beta} + \mathbf{w}_{it}'\boldsymbol{\alpha}_i + \varepsilon_{it}, \tag{4.16}$$

where $\mathbf{z}_{it}$ includes an intercept, $\mathbf{w}_{it}$ is a vector of observable characteristics, $\boldsymbol{\beta}$ is a **fixed** parameter vector and $\boldsymbol{\alpha}_i$ is a **random zero-mean vector**.

The name "mixed" refers to that the model consists of both **fixed**, $\boldsymbol{\beta}$, and **random**, $\boldsymbol{\alpha}_i$, parameters.

The random effects model is a special case of the mixed model with $\mathbf{w}_{it}'\boldsymbol{\alpha}_i = \alpha_i$ (random intercepts). In contrast, the **random coefficients model** specifies

$$y_{it} = \mathbf{z}_{it}'\boldsymbol{\beta}_i + \varepsilon_{it}, \tag{4.17}$$

so that the parameter vector now varies across individuals according to $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$, equivalent to (4.16) with $\mathbf{w}_{it} = \mathbf{z}_{it}$.

Covariates in the $\mathbf{z}_{it}$ vector capture differences in the conditional **mean** while the $\mathbf{w}_{it}$ vector captures differences in the conditional **variance**.

Mixed Linear Models: Estimation

Model (4.17) has a **deterministic**, $\mathbf{z}_{it}'\boldsymbol{\beta}$, and a **random**, $\mathbf{w}_{it}'\boldsymbol{\alpha}_i + \varepsilon_{it}$, part. The goal is to estimate the vector $\boldsymbol{\beta}$ and the VCV matrix, $\boldsymbol{\Omega}(\boldsymbol{\alpha}_i, \varepsilon_{it})$.

Under the RE assumption of independence between regressors and the random components, POLS of y_{it} on $\mathbf{x}_{it}$ provide consistent estimates of $\boldsymbol{\beta}$.

Since structure has been put on the error term $(\mathbf{w}_{it}'\boldsymbol{\alpha}_i + \varepsilon_{it})$ FGLS estimation will be more efficient, however. The variance components can be obtained using different **maximum likelihood methods**. Alternatively, such models are estimated by **bayesian methods**.

We may obtain **predictors** for the random slopes and random intercepts. They cannot be consistently estimated but they may still be interpreted in a **best predictor** sense. To obtain such predictors Bayesian Analysis, e.g. **Markov-Chain Monte Carlo-Methods (MCMC)**, is convinient.

Random Effects Models and Best Predictors

Example 4.4 (Paul (SJOE, 2016))

The Scandinavian
Journal of Economics

Scand. J. of Economics 00(00), 1–12, 2016
DOI: 10.1111/sjoe.12157

Is There a Causal Effect of Working Part-Time on Current and Future Wages?*

Marie Paul

University of Duisburg-Essen, Duisburg DE-47057, Germany
marie.paul@uni-due.de

Abstract

In this paper, I study the causal effects of part-time work on current and future wages. To estimate these effects, I use a random effects model with a wage equation capturing the employment history and a dynamic multinomial probit component for the choice of employment status. Exclusion restrictions from the institutional context are exploited to support identification. The results suggest that working part-time with few hours has a large causal effect on current wages, but more extensive part-time work does not reduce current wages. However, both types of part-time work lead to negative long-term wage effects.

Keywords: Female employment patterns; Markov chain Monte Carlo methods; part-time employment; random effects models

JEL classification: *C*11; *C*33; *J*16; *J*24; *J*31

Random Effects Models and Best Predictors

Example 4.5 (Paul, SJOE, 2016)

$$\ln W_{it} = \ln W_{it}^{*} \cdot \mathbb{1}(E_{it} \neq 3);$$

$$\ln W_{it}^{*} = \beta_0^W + \beta_1^W PTS_{it} + \beta_2^W PTL_{it} + H_{it}^{W'} \beta_3^W + x_{it}^{W'} \beta_4^W + \alpha_i^W + \varepsilon_{it}^W.$$

$$U_{itj} = \beta_{0j}^E + H_{it}^{E'} \beta_{1j}^E + x_{it}^{E'} \beta_{2j}^E + z_{itj}^{E'} \beta_{3j}^E + \alpha_{ij}^E + \varepsilon_{itj}^E.$$

$$E_{it} = h(Y_{it}) \equiv \begin{cases} 0 & \text{if} \quad \max(Y_{it}) \leq 0 \\ j & \text{if} \quad \max(Y_{it}) = U_{itj} > 0. \end{cases}$$

Random Slopes

Example 4.6 (Your own research)

Would it be interesting to allow for random slopes in your research?