Econometrics II

Chapter 3
Panel data models I

Marie Paul

University of Duisburg-Essen
Ruhr Graduate School in Economics

Summer Semester 2020

Outline of lecture

3. Panel data models I
    - Motivation
    - Basic panel data concepts and models
    - Basic panel data estimators
    - Application: Comparison of POLS, RE, FE

Self-study plan for Lecture III

- This lecture deals with basic linear panel data models and estimators. The focus is on comparison and differentiation of different estimators and on some specific aspects that have practical implications but are sometimes not discussed in undergraduate courses.
- For our third virtual meeting, please work through all slides of Lecture III.
- The slides mostly follow Cameron and Trivedi Chapter 21 (Linear Panel Models: Basics). Please read the text as far as it relates to the slides.
- In general, Wooldridge's textbooks are great on panel data models. You may want to complement your reading of Cameron and Trivedi with either the intuitive, but often revealing text on POLS, FE, and RE in "Introductory Econometrics" or the more formal exposition in his graduate text book ("Econometrics Analysis of...").
- You do not need to read the paper from the application (Fernandez-Kranz et al.), because we only cover a special case from the sensitivity analysis of this paper.
- As usual, please prepare the "your own research" questions and collect questions and topics to discuss them in our virtual meeting.

Outline

3. Panel data models I
   - Motivation
   - Basic panel data concepts and models
   - Basic panel data estimators
   - Application: Comparison of POLS, RE, FE

Definition

Definition 3.1 (Panel data)

Panel (longitudinal) data are repeated **microeconomic** observations on the same cross-section observed for several **time periods**. The model

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}' \boldsymbol{\beta}_{it} + u_{it} \tag{3.1}$$

is a general panel data specification for cross section unit i in time period t. Focus on short panels.

There are several advantages with panel data compared to cross-sectional data:

1. **Increased precision: Pooling** the observations for each individual over time results in more observations and smaller standard errors.
2. **Omitted variable bias:** By estimating **fixed** or **random effects** models one can avoid problems of **unobserved heterogeneity**.
3. **Dynamics:** Panel data allow for analyzing **dynamic behavior** of individuals and **dynamic models** using lagged outcome variables.

Definition

Example 3.2 (Effect of job training on wages)

$$log(wage)_{it} = \alpha_{it} + \mathbf{x}'_{it}\boldsymbol{\beta}_{it} + u_{it} \tag{3.2}$$

Yearly data on employee's hourly wages in a firm. Here $\mathbf{x}_{it}$ includes dummies if a person i participates in 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.

Motivation: Omitted variables revisited

Panel data can assist in identification when **unobserved individual heterogeneity** poses a problem. To see this, consider the linear cross-section model

$$\mathrm{E}[y_i|\mathbf{x}_i, \alpha_i] = \alpha_i + \mathbf{x}_i'\boldsymbol{\beta}, \tag{3.3}$$

Note: Wooldridge (2010) uses c to characterize an **unobserved random variable** so as to clearly distinguish it from the **parameter** α.

Interest is in the vector $\boldsymbol{\beta}$. Whether the random variables α_i pose a problem or not in the identification of $\boldsymbol{\beta}$ depends on their relation to $\mathbf{x}_i$.

Case 1. $\mathrm{Cov}(\mathbf{x}_i, \alpha_i) = \mathbf{0}$: α_i is just another unobserved factor affecting y but **do not interfere** with consistent estimation of $\boldsymbol{\beta}$.

Case 2. $\mathrm{Cov}(\mathbf{x}_i, \alpha_i) \neq \mathbf{0}$: Will generate endogeneity between regressors and the error term and estimation of $\boldsymbol{\beta}$ will be **inconsistent**.

Hence, the choice of estimator of $\boldsymbol{\beta}$ will critically depend on what is assumed about the relationship between $\mathbf{x}$ and α_i, e.g., ability and returns to education.

Motivation: Omitted variables revisited

Unobserved endogeneity can be tackled by the IV approach. However, finding suitable instruments is difficult and subject to other limitations.

Instead, exploiting the **panel data dimension** it is possible to adjust for this type of endogeneity to obtain consistent estimation of $\boldsymbol{\beta}$.

Consider the extension of (3.3) by adding a time subscript

$$\mathrm{E}[y_{it}|\mathbf{x}_{it}, \alpha_i] = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \quad t = 1, 2. \tag{3.4}$$

Here, y_{it} and $\mathbf{x}_{it}$ are observed in two **distinct** time periods $t = 1, 2$ whereas α is a **time constant** and **individual-specific** unobserved effect; i.e., $\alpha_{i1} = \alpha_{i2}$.

In this sense, α_i is interpreted as a set of given features of a person or a firm, such as **cognitive ability** or **managerial quality**.

The **time invariance** of the **idiosyncratic variable**, α_i, is a crucial assumption for the models discussed in this lecture.

Motivation: Omitted variables revisited

To see this, write model (3.4) in error form as

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \overbrace{\alpha_i + \varepsilon_{it}}^{u_{it}}, \tag{3.5}$$

where $u_{it} = \alpha_i + \varepsilon_{it}$ is a composite error term and we assume **strict exogeneity**,

$$\mathsf{E}[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}] = 0 \Rightarrow \mathsf{E}[\mathbf{x}'_{is}\varepsilon_{it}] = \mathbf{0}, \quad s, t = 1, 2, \tag{3.6}$$

i.e., that ε_{it} has mean zero conditional on **past, current and future** values of $\mathbf{x}$.

If we additionally assume that

$$\mathsf{E}[\mathbf{x}'_{it}\alpha_i] = \mathbf{0}, \tag{3.7}$$

we could simply **pool** the data over time and apply OLS to consistently estimate $\boldsymbol{\beta}$. However, if (3.7) does not hold, OLS will be inconsistent.

If we instead **differentiate** (3.5) so that

$$\Delta y_i = \boldsymbol{\Delta}\mathbf{x}'_i\boldsymbol{\beta} + \Delta\varepsilon_i, \tag{3.8}$$

where $\Delta q_i = q_{i2} - q_{i1}$ is generic for $q = (y, \mathbf{x}, \varepsilon)$, we obtain a linear model in the **differences** where α_i has been dropped out.

Motivation: Omitted variables revisited

Since (3.8) is now a cross-section, OLS assumptions for the **differentiated** model apply for consistency of $\boldsymbol{\beta}$. Thus,

OLS.1. $\mathrm{E}[\boldsymbol{\Delta}\mathbf{x}_i'\Delta\varepsilon_i] = \mathbf{0}$. (orthogonality)

OLS.2. rank $\mathrm{E}[\boldsymbol{\Delta}\mathbf{x}_i'\boldsymbol{\Delta}\mathbf{x}_i] = K$. (rank)

OLS.1 can be written

$$\mathrm{E}[\mathbf{x}_{i2}'\varepsilon_{i2}] + \mathrm{E}[\mathbf{x}_{i1}'\varepsilon_{i1}] - \mathrm{E}[\mathbf{x}_{i1}'\varepsilon_{i2}] - \mathrm{E}[\mathbf{x}_{i2}'\varepsilon_{i1}] = \mathbf{0}, \tag{3.9}$$

which holds by assumption (3.6). For the two last terms we need the stronger assumption of **strict exogeneity** of $\mathbf{x}_{it}$ and ε_{it}.

OLS.2 is also important here since all elements of $\mathbf{x}$ constant over time for all individuals will **cancel out** in the differentiation, i.e.,

$$\Delta\mathbf{x}_i = \mathbf{x}_{i2} - \mathbf{x}_{i1} = 0, \tag{3.10}$$

implying that we will not be able to identify the **partial effect** of these factors.

This is intuitive as the impact on y from variation in these variables cannot be **distinguished** from the variation in α_i.

Outline

3. Panel data models I
   - Motivation
   - Basic panel data concepts and models
   - Basic panel data estimators
   - Application: Comparison of POLS, RE, FE

General linear model for panel data

Focus on a **short** (i.e. N goes to infinity, but T does not) and **balanced** panel. Most results apply also to unbalanced panels and Stata can **handle unbalanced panels** for estimation.

The **general panel data model** setup allows the intercept α and slope β coefficients to vary over both individuals and time

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{3.11}$$

This model is not estimable since there are more parameters than observations. Restrictions are needed on u_{it} and how α_{it} and β_{it} varies with i and t.

General linear model for panel data

Common **panel data model** specifications are:

- **Pooled model**: Specifies **constant coefficients** as in cross section.

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it} \tag{3.12}$$

- **Individual-specific effects model**: Permits **intercepts** but not **slope parameters** to vary across individuals.

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \tag{3.13}$$

  where ε_{it} is iid over i and t. There may be time dummies γ_t. They are included in $\mathbf{x}_{it}$

General linear model for panel data

Example 3.3 (Effect of job training on wages (contin.))

In example 3.2 the individual-specific effects model allows ability and motivation to differ across workers also conditional on covariates like education:

$$log(wage)_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \tag{3.14}$$

- It assumes that ability and motivation are time constant, time-variant aspects of it will be in the error term.
- The individual-specific effects model assumes that slopes are the same for individuals, e.g. the coefficient for college degree is the same for all workers.
- This assumption could be relaxed by allowing for individual specific slopes. More common are interaction effects, e.g. the effect of college degree is allowed to vary by occupation.

Strict Exogeneity

Throughout this chapter we make the assumption of strong exogeneity or strict exogeneity (conditional on the unobserved effect):

$$\mathrm{E}[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = 0, \quad t = 1, ..., T. \tag{3.15}$$

i.e., that ε_{it} has mean zero conditional on **past, current and future** values of $\mathbf{x}$. or, together with the functional form assumption:

$$\mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = \mathrm{E}[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}, \quad t = 1, ..., T. \tag{3.16}$$

It means that, once $\mathbf{x}_{it}$ and α_i are controlled for, $\mathbf{x}_{is}$ has no partial effect on y_{it} for $s \neq t$.
Strict exogeneity implies that the explanatory variable in each time period are uncorrelated with the idiosyncratic error in each time period:

$$\mathrm{E}[\mathbf{x}_{is}', \varepsilon_{is}] = \mathbf{0}, \quad s, t = 1, ..., T. \tag{3.17}$$

Strict Exogeneity

Example 3.4 (Effect of job training on wages (contin.))

- Strict exogeneity implies that ε_{it} is uncorrelated with **future program participation**. An exogenous chock in t (e.g. new software introduced) may lower the wage of those who do not master it in t and make the managers assign training to those in $t+1$. Such **feed-back effects** are ruled out by the strict exogeneity assumption.
- Strict exogeneity also implies that ε_{it} is uncorrelated with past program participation. This appears if a long-lasting effect of the program is captured in a later ε_{it}. This problem can be solved by **adding enough lags** of program participation to the model.
- Strict exogeneity **fails** in a dynamic model with **lagged dependent variables** (to be discussed in Chapter 4), thus we cannot add $log(wage)_{it-1}$.

Between and within variation

Between and **within variation** refers to empirical variation across (y_{it} and y_{jt}) and within cross section units over time (y_{it} and y_{is}), respectively.

The **between model** is a special case of the pooled model where we average each component in (3.34) across all time periods,

$$\begin{aligned} \bar{y}_i &= \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i \\ &= \alpha + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + (\alpha_i - \alpha + \bar{\varepsilon}_i), \end{aligned} \tag{3.18}$$

where $\bar{q}_i = T^{-1} \sum_t q_{it}$ is generic for $q = (y, \mathbf{x}, \varepsilon)$.

The **within model** measures the relation between individual-specific **deviations** of regressors and outcomes from their time-averaged values

$$y_{it} - \bar{y}_i = (\alpha_i - \alpha_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i), \tag{3.19}$$

where (3.18) has been subtracted from (3.34).

Keep in mind: While the **within variation** is removed from **(3.18)**, the **between variation** is removed from **(3.19)**.

Summarizing data variation

For panel data it is useful to know whether variability is mostly **across individuals** or **across time**.

Total variation in x_{it} can be decomposed as

$$\begin{aligned}\underbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2}_{TSS} &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left[(x_{it}-\bar{x}_i)+(\bar{x}_i-\bar{x})\right]^2 \\ &= \underbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}_{WSS}+\underbrace{\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2}_{BSS}.\end{aligned} \tag{3.20}$$

The **total** sum of squares (TSS) equals the **within** sum of squares (WSS) plus the **between** sum of squares (BSS) with sample variances

$$s_W^2 = \frac{1}{NT-N}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2 \quad \text{and} \quad s_B^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2 \tag{3.21}$$

Identification in random effects models

The identification problem is that we cannot estimate the conditional mean

$$E[y_{it}|\alpha_i, \mathbf{x}_{it}] = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} \tag{3.22}$$

because of the presence of α_i.

Both the **fixed effects (FE)** and the **random effects (RE)** models takes α_i into account but vary with respect to its assumed relationship with $\mathbf{x}_{it}$.

One variant of the model 3.17 treats α_i **as unobserved random variables potentially correlated with the observables x** . Called fixed-effects (FE) model.

One other variant of the model 3.17 treats α_i **as unobserved random variables distributed independently** of the regressors. Called random effects (RE) model. Usually here additional assumptions: $\alpha_i \sim [\alpha, \sigma_\alpha^2]$ and $\varepsilon_{it} \sim [0, \sigma_\varepsilon^2]$

Identification in random effects models

The individual effect is a random variable in both fixed and random effects models. Use the following notation:

$$y_{it} = c_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3.23}$$

Both models assume that:

$$\mathsf{E}[y_{it}|c_i, \mathbf{x}_{it}] = c_i + \mathbf{x}'_{it}\boldsymbol{\beta} \tag{3.24}$$

Take the expectation with respect to c_i to eliminate c_i:

$$\mathsf{E}[y_{it}|\mathbf{x}_{it}] = E[c_i|\mathbf{x}_{it}] + \mathbf{x}'_{it}\boldsymbol{\beta} \tag{3.25}$$

For the RE model it is assumed that $E[c_i|\mathbf{x}_{it}] = \alpha$ and thus constant (in the estimation equation α_i is then part of the error term), so $\mathsf{E}[y_{it}|\mathbf{x}_{it}] = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta}$ and hence $\mathsf{E}[y_{it}|\mathbf{x}_{it}]$ **the CEF can be identified**.

Identification in fixed effects models

In the **FE model**, $\mathrm{E}[c_i|\mathbf{x}_{it}]$ varies with $\mathbf{x}_{it}$ and hence it is impossible to identify the CEF $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$, because we cannot put c_i in the error term as it correlated to regressors and c_i cannot be consistently estimated in short panels.

Remarkably, for the **linear model** there are several ways to consistently estimate β despite the presence of these parameters.

So we can nevertheless **consistently estimate** β in the FE model. It is thus possible to identify the **marginal effect** β for time-variable regressors even though this involves the CEF which is not identified.

Panel data models

Example 3.5 (Your own research)

Are panel data models used in your research area? Individual specific effects models or even more flexible models? Are there examples for RE and FE models?

Outline

3 Panel data models I
- Motivation
- Basic panel data concepts and models
- Basic panel data estimators
- Application: Comparison of POLS, RE, FE

Panel data estimators: Overview

Panel data estimators vary in to which extent **cross-sectional** and/or **time-series** variation in the data are used in the estimation of β.

The panel data estimators can be written on a **transformed OLS form**.

Table 3.1. Linear Panel Data Estimators of $\boldsymbol{\beta}$

Estimator		Definition	Transformation
Pooled OLS	$\widehat{\boldsymbol{\beta}}_{POLS}$	$= \left[\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right]^{-1} \sum_{i=1}^{N} \mathbf{X}_i'\mathbf{y}_i$	$\mathbf{X}_i = \sum_t \mathbf{x}_{it}$
Between	$\widehat{\boldsymbol{\beta}}_{B}$	$= \left[\sum_{i=1}^{N} \bar{\mathbf{X}}_i'\bar{\mathbf{X}}_i\right]^{-1} \sum_{i=1}^{N} \bar{\mathbf{X}}_i'\bar{\mathbf{y}}_i$	$\bar{\mathbf{X}}_i = T^{-1} \sum_t \mathbf{x}_{it}$
Within/LSDV	$\widehat{\boldsymbol{\beta}}_{W}$	$= \left[\sum_{i=1}^{N} (\ddot{\mathbf{X}}_i)'(\ddot{\mathbf{X}}_i)\right]^{-1} \sum_{i=1}^{N} (\ddot{\mathbf{X}}_i)'(\ddot{\mathbf{y}}_i)$	$\ddot{\mathbf{X}}_i = \sum_t (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$
First Differences	$\widehat{\boldsymbol{\beta}}_{FD}$	$= \left[\sum_{i=1}^{N} (\Delta\mathbf{X}_i)'(\Delta\mathbf{X}_i)\right]^{-1} \sum_{i=1}^{N} (\Delta\mathbf{X}_i)'(\Delta\mathbf{y}_i)$	$\Delta\mathbf{X}_i = \sum_t (\mathbf{x}_{it} - \mathbf{x}_{it-1})$
Random Effects	$\widehat{\boldsymbol{\beta}}_{RE}$	$= \left[\sum_{i=1}^{N} (\tilde{\mathbf{X}}_i)'(\tilde{\mathbf{X}}_i)\right]^{-1} \sum_{i=1}^{N} (\tilde{\mathbf{X}}_i)'(\tilde{\mathbf{y}}_i)$	$\tilde{\mathbf{X}}_i = \sum_t (\mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i)$

where $\lambda = 1 - \sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$ is a quasi-time demeaning parameter. Note that

- $\widehat{\boldsymbol{\beta}}_{RE} \rightarrow \widehat{\boldsymbol{\beta}}_{POLS}$ when $\lambda \rightarrow 0$.
- $\widehat{\boldsymbol{\beta}}_{RE} \rightarrow \widehat{\boldsymbol{\beta}}_{W}$ when $\lambda \rightarrow 1$ (or $T \rightarrow \infty$).

Pooled OLS (POLS)

The **pooled OLS estimator**, $\boldsymbol{\beta}_{POLS}$, is obtained by stacking the data over i and t into a long regression with NT observations and using OLS.

If we have the Pooled Model

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}. \tag{3.26}$$

we consistently estimated by $\widehat{\boldsymbol{\beta}}_{POLS}$ under the assumptions

POLS.1. $\mathrm{E}[\mathbf{x}_{it}' u_{it}] = \mathbf{0}$, $i = 1, 2, ..., N$, $t = 1, 2, ..., T$.

POLS.2. rank $\left[\sum_{t=1}^{T} \mathrm{E}[\mathbf{x}_{it}' \mathbf{x}_{it}]\right] = K$.

Remember that this model does not include an individual specific effect.

Under POLS.1 and POLS.2 consistency of $\widehat{\boldsymbol{\beta}}_{POLS}$ requires that either $N \to \infty$ or $T \to \infty$ (compared to $N \to \infty$ in standard OLS).

However, $\widehat{\boldsymbol{\beta}}_{POLS}$ will be inefficient and a standard variance matrix is wrong unless additional assumptions of **homoscedasticity** and no **serial correlation** are maintained

POLS.3. (a) $\mathrm{E}[u_{it}^2 \mathbf{x}_{it}' \mathbf{x}_{it}] = \sigma^2 \mathrm{E}[\mathbf{x}_{it}' \mathbf{x}_{it}]$; (b) $\mathrm{E}[u_{it} u_{is} \mathbf{x}_{it}' \mathbf{x}_{is}] = \mathbf{0}$, $t \neq s$.

Panel robust variance estimators

Assumption POLS.3 is a **strong restriction** in panel data models as it implies that errors are uncorrelated for each **individual over time** ($\mathbf{w}_{it} = [1, \mathbf{x}_{it}]$).

$$\mathbf{\Omega}^{\mathsf{POLS}} = \mathsf{E}\left[\sum_{i}^{N} \mathbf{u}_i \mathbf{u}_i' | \mathbf{W}_i\right] = \sigma^2 \mathbf{I}_{NT} \equiv \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} \qquad (3.27)$$

POLS.3 is motivated by iid sampling. But errors for a given individual i are almost certainly **positively correlated over t**, even after inclusion of regressors $Cor[u_{it}, u_{is}]$ is likely to remain high. This must be accounted for **not to understate the variance**.

The usual OLS output treats each of the T years as independent pieces of information, but information content is less than this given the positive error correlations. This can lead to large **overstatement of estimator precision**.

Panel robust variance estimators

If POLS.3 fails, a **panel robust estimate** of the asymptotic variance matrix is one that safeguards against both **heteroscedasticity** and **serial correlation**

$$\widehat{\mathrm{V}}[\boldsymbol{\beta}_{POLS}] = \left[\sum_{i}^{N} \mathbf{W}_i' \mathbf{W}_i\right]^{-1} \sum_{i}^{N} \mathbf{W}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{W}_i \left[\sum_{i}^{N} \mathbf{W}_i' \mathbf{W}_i\right]^{-1}, \qquad (3.28)$$

where $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i \widehat{\boldsymbol{\beta}}_{POLS}$. It is consistent in **short panels** $(N \to \infty, T \nrightarrow \infty)$ since the dimension of $\mathbf{u}_i$ is then a finite $T \times T$ matrix.

Panel robust variance estimators in general

In all panel data estimators listed in Table 3.1, **heteroscedasticity** and **serial correlation** may generate incorrect statistical inference.

The **panel-robust asymptotic variance matrix** of the pooled OLS estimator controls for both these issues. Write the **generic** OLS model as

$$\check{\mathbf{y}} = \breve{\mathbf{W}}'\boldsymbol{\theta} + \check{\mathbf{u}}, \tag{3.29}$$

where the checkmark is generic for the FE, RE and FD **transformations**, $\boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}]$ and $\mathbf{W} = [1, \mathbf{X}]$, respectively.

The OLS estimator is

$$\hat{\boldsymbol{\theta}}_{OLS} = \left[\breve{\mathbf{W}}'\breve{\mathbf{W}}\right]^{-1}\breve{\mathbf{W}}'\check{\mathbf{y}}, \tag{3.30}$$

with the panel-robust asymptotic variance **estimator**. In Stata: vce(cluster panelvar) in some commands equivalent to vce(robust).

$$\hat{\mathrm{V}}[\boldsymbol{\theta}_{OLS}] = \left[\breve{\mathbf{W}}'\breve{\mathbf{W}}\right]^{-1}\breve{\mathbf{W}}'\hat{\mathbf{u}}\hat{\mathbf{u}}'\breve{\mathbf{W}}\left[\breve{\mathbf{W}}'\breve{\mathbf{W}}\right]^{-1}, \tag{3.31}$$

where $\hat{\mathbf{u}} = \check{\hat{\mathbf{u}}} = \check{\mathbf{y}} - \breve{\mathbf{W}}\hat{\boldsymbol{\beta}}$.

Inconsistency of the pooled OLS model

The **pooled model** is inconsistent if the true model is the **individual-specific effects model** and **the individual specific effect is correlated with regressors**:

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \tag{3.32}$$

This can be seen by rewriting the model with a common intercept

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \overbrace{(\alpha_i - \alpha + \varepsilon_{it})}^{u_{it}}. \tag{3.33}$$

Here, the OLS assumption that $\mathrm{E}[\mathbf{x}_{it}u_{it}] = 0$ will not hold if the individual specific effect α_i is correlated with the regressors.

In summary:

- Pooled OLS is appropriate if the **constant coefficient** model is appropriate.
- Compared to ordinary OLS it will give improved estimation **precision**.
- Even if consistent, **standard errors** should be **corrected** for panel data.
- Pooled OLS is also consistent for the individual specific effects model if **POLS.1 holds for the comprehensive error term**. But in this case more efficient estimators exist.

The individual-specific effects model

We turn to estimators for the individual-specific effects model

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \tag{3.34}$$

where α_i are random variables capturing **time-constant unobserved heterogeneity**. This model nests the pooled OLS model.

Furthermore now we need to assume that $\mathbf{x}_{it}$ is **strictly exogenous** from (3.6) implying that $\mathrm{E}[\mathbf{x}_{is}'\varepsilon_{it}] = \mathbf{0}$ for $s, t = 1, ..., T$.

Four ways to obtain a consistent **estimates of** β under strict exogeneity and varying further assumptions:

1. The Individual and time dummy variables model.
2. The Random effects (RE) model.
3. The Fixed effects (FE) model.
4. The First differences (FD) model.

The dummy variables estimator

We now focus on the case where pooled OLS is inconsistent, the fixed-effects model. That is, when $\mathrm{E}[\mathbf{x}'_{it}\alpha_i] \neq \mathbf{0}$ so that the unobserved **fixed effect** α_i must be eliminated.

The **individual and time dummy variables model** is the representation of (3.34) with **dummy variables** for each individual and time period

$$\begin{aligned} y_{it} &= (\alpha_i + \gamma_t) + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ &= \sum_{j=1}^{N} \alpha_j d_{j,it} + \sum_{s=2}^{T} \gamma_s d_{s,it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \end{aligned} \tag{3.35}$$

where $d_m = \mathbf{1}[m = n]$ for $[m, n] = [(i, t), (j, s)]$ and equivalent to model (3.34) with γ_1 subsumed into the regressor vector $\mathbf{x}_{it}$.

The estimator of $\boldsymbol{\beta}$ in (3.35), $\widehat{\boldsymbol{\beta}}_{LSDV}$, is consistent if $N \to \infty$ and $T \to \infty$.

The **incidental parameters problem** implies inconsistency of $\widehat{\boldsymbol{\beta}}_{LSDV}$ when as in our case of short panels only $N \to \infty$ since there are $N + (T - 1) + K$ parameters with only NT observations. Note that remarkebly β can still be consistently estimated.

The fixed effects estimator

In practice, as most panel data have large N it is **not recommended** to estimate the individual-specific effects model with dummy variables.

The standard way to obtain the β is to use the **fixed-effects estimator (within-estimator)**.

The fixed-effects estimator measures the association between individual-specific deviations of regressors from their **time-averaged values** and individual-specific deviations of the dependent variable form its time-averages value.

The fixed effects estimator

The **fixed effects (within) estimator** is based on the **within model** (3.19), which is simply the **demeaned** version of (3.34)

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}'\boldsymbol{\beta} + \ddot{\varepsilon}_{it}, \tag{3.36}$$

where $\ddot{q}_{it} = q_{it} - \bar{q}_i$ is generic for $q = (y, \mathbf{x}, \varepsilon)$.

The assumptions required for consistency of the fixed effects estimator are

FE.1. $\mathsf{E}[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = 0, \quad t = 1, ..., T,$

FE.2. rank $[\sum_{t=1}^{T} \mathsf{E}[\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}]] = K.$

Under FE.1 $\mathsf{E}[\ddot{\varepsilon}_{it}|\ddot{\mathbf{x}}_{it}] = 0$ and given FE.2 the within estimator of $\boldsymbol{\beta}$ can be consistently estimated by pooled OLS as

$$\widehat{\boldsymbol{\beta}}_W = \left[\sum_{i=1}^{N} (\ddot{\mathbf{X}}_i)'(\ddot{\mathbf{X}}_i)\right]^{-1} \sum_{i=1}^{N} (\ddot{\mathbf{X}}_i)'(\ddot{\mathbf{y}}_i). \tag{3.37}$$

The fixed effects estimator

The **individual fixed effect** can be estimated by:

$$\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\boldsymbol{\beta}}_w \tag{3.38}$$

The estimate $\hat{\alpha}_i$ is unbiased for α_i but **not consistent** unless $T \to \infty$ since it involves averaging of T observations.

The remarkable results that estimation of $\boldsymbol{\beta}$ is still consistent does not carry over to more complicated FE models as **nonlinear models**.
A major limitation of this estimator is that coefficients of **time-invariant regressors are not identified** in the within model.

The fixed effects estimator

The within estimator is **efficient** under the assumption of iid errors

FE.3. $\mathrm{E}[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \mathbf{x}_i, \alpha_i]] = \sigma_u^2 \mathbf{I}_T$

which leads to the estimator for the asymptotic variance matrix

$$\hat{\mathrm{V}}[\hat{\boldsymbol{\beta}}_W] = \sigma_\varepsilon^2 \left[\sum_{i=1}^{N} (\ddot{\mathbf{X}}_i)' (\ddot{\mathbf{X}}_i) \right]^{-1} \tag{3.39}$$

A preferred estimator permits arbitrary autocorrelations for the ε_{it} and arbitrary heteroskedasticity.

The first differences estimator

The **first differences model** is obtained by subtracting the **one period lagged model**

$$y_{it-1} = \alpha_i + \mathbf{x}'_{it-1}\boldsymbol{\beta} + \varepsilon_{it-1}, \quad t = 1, ..., T, \tag{3.40}$$

from (3.34).

This yields

$$y_{it} - y_{it-1} = (\mathbf{x}_{it} - \mathbf{x}_{it-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{it-1}), \quad t = 1, ..., T, \tag{3.41}$$

which can be equivalently written

$$\boldsymbol{\Delta}\mathbf{y}_i = \boldsymbol{\Delta}\mathbf{X}'_i\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\varepsilon}_i, \tag{3.42}$$

where $\boldsymbol{\Delta}\mathbf{q}_i = \sum_t(\mathbf{q}_{it} - \mathbf{q}_{it-1})$ is a generic $T-1$ vector for $\mathbf{q} = (\mathbf{y}, \mathbf{X}, \boldsymbol{\varepsilon})$.

Note that the **first period is dropped** in this model so that there remains only $N(T-1)$ observations.

The first difference estimator

The **first difference (FD) estimator** is the **pooled OLS estimator** of the first differenced model. Hence, estimating

$$\boldsymbol{\Delta}\mathbf{y}_i = \boldsymbol{\Delta}\mathbf{X}_i'\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\varepsilon}_i, \tag{3.43}$$

with OLS yields

$$\widehat{\boldsymbol{\beta}}_{FD} = \left[\sum_{i=1}^{N}(\boldsymbol{\Delta}\mathbf{X}_i)'(\boldsymbol{\Delta}\mathbf{X}_i)\right]^{-1}\sum_{i=1}^{N}(\boldsymbol{\Delta}\mathbf{X}_i)'(\boldsymbol{\Delta}\mathbf{y}_i). \tag{3.44}$$

The FD estimator is consistent under the assumptions (may be somewhat relaxed)

FD.1. $\mathrm{E}[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = 0, \quad t = 1, ..., T$

FD.2. rank $\left[\sum_{t=2}^{T} \mathrm{E}[\boldsymbol{\Delta}\mathbf{x}_{it}'\boldsymbol{\Delta}\mathbf{x}_{it}]\right] = K$

It is also efficient under the following assumption

FD.3. (a) $\mathrm{E}[e_{it}^2\mathbf{x}_{it}'\mathbf{x}_{it}] = \sigma^2\mathrm{E}[\mathbf{x}_{it}'\mathbf{x}_{it}]$; (b) $\mathrm{E}[e_{it}e_{is}\mathbf{x}_{it}'\mathbf{x}_{is}] = \mathbf{0}, \ t \neq s$

where $e = \Delta\varepsilon_{it}$, yielding the error structure $\varepsilon_{it} = \varepsilon_{i,t-1} + e_{it}$. Otherwise a robust variance matrix can be computed along the usual lines.

Hence, while assumption FE.3 maintains that errors are **serially uncorrelated**, assumption FD.3 maintains that errors follow a **random walk**.

Fixed-effects of First Differences?

In practice results are **often very similar**.

If T = 2 they are identical.

FE is more efficient when the errors are **serially uncorrelated** and FD when they follow a **random walk**. Often we have something in between.

The estimators differ in how severe violations of different assumptions are, see Wooldridge (2010) for a discussion.

The random effects estimator

The RE model is a special case of the pooled model with the unobserved effect **subsumed** in the **error term**.

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \overbrace{(\alpha_i + \varepsilon_{it})}^{u_{it}}, \tag{3.45}$$

The **composite errors** u_{it} will be **serially correlated** since

$$\mathrm{Cov}[u_{it}, u_{is}] = \mathrm{E}[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})] = \begin{cases} \sigma_\alpha^2, & t \neq s \\ \sigma_\alpha^2 + \sigma_\varepsilon^2, & t = s \end{cases} \tag{3.46}$$

so pooled OLS will be **consistent but inefficient** under the RE model.

The random effects estimator

The **random effects estimator** is typically estimated with **Feasible GLS** under the random effects assumptions ($\mathbf{w}_{it} = [1, \mathbf{x}_{it}]$)

RE.1. (a) $\mathsf{E}[\varepsilon_{it}|\alpha_i, \mathbf{w}_{i1}, ..., \mathbf{w}_{iT}] = 0, \quad t = 1, ..., T$
(b) $\mathsf{E}[\alpha_i|\mathbf{w}_{it}] = \mathsf{E}[\alpha_i] = 0$

RE.2. rank $\mathsf{E}[\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{W}_i]] = K$

RE.3. (a) $\mathsf{E}[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{w}_i, \alpha_i]] = \sigma_\varepsilon^2 \mathbf{I}_T$
(b) $\mathsf{E}[\alpha_i^2|\mathbf{w}_i] = \sigma_\alpha^2$

Assumption RE.3 ascertain that errors have the **random effects structure**

$$\boldsymbol{\Omega}_{\mathsf{RE}} = \mathsf{E}\left[\sum_i^N \mathbf{u}_i\mathbf{u}_i'|\mathbf{W}_i\right] = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix} \tag{3.47}$$

Note that the correlation between errors does not vary over time in the RE model: they are **equicorrelated**.

The random effects estimator

The **FGLS estimator** of the RE model is the **random effects estimator**

$$\hat{\beta}_{\text{RE}} = \left[\mathbf{W}'\hat{\Omega}_{\text{RE}}^{-1}\mathbf{W}\right]^{-1}\mathbf{W}'\hat{\Omega}_{\text{RE}}^{-1}\mathbf{y}, \tag{3.48}$$

with $\hat{\Omega}_{\text{RE}} = \hat{\sigma}_{\varepsilon}^2\mathbf{I}_T + \hat{\sigma}_{\alpha}^2\mathbf{j}_T\mathbf{j}_T'$, where $\mathbf{j}_T' = (1, \ldots, 1)^T$.

The asymptotic variance estimator is equal to

$$\hat{\text{V}}[\hat{\beta}_{\text{RE}}] = \left[\sum_{i=1}^{N}\mathbf{W}_i'\hat{\Omega}_{\text{RE}}^{-1}\mathbf{W}_i\right]^{-1} \tag{3.49}$$

where consistent estimates $\hat{\sigma}_{\alpha}^2$ and $\hat{\sigma}_{\varepsilon}^2$ can be obtained from a between and a within regression of y on $\bar{\mathbf{w}}$, respectively.

More robust variance estimation could be obtained by a more **general FGLS analysis** or with **panel-robust estimation**.

Example 3.6 below shows how GLS estimation of (3.45) is equivalent to OLS estimation of the transformed FE model in Table 3.1.

Random and fixed effects estimators

Example 3.6 (Relationship between $\widehat{\beta}_{RE}$ and $\widehat{\beta}_{FE}$)

Rewrite $\mathbf{\Omega}_{\mathsf{RE}}$ as

$$\begin{aligned}\mathbf{\Omega}_{\mathsf{RE}} &= \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{j}_T \mathbf{j}_T' = \sigma_\varepsilon^2 \mathbf{I}_T + T\sigma_\alpha^2 \mathbf{j}_T (\mathbf{j}_T \mathbf{j}_T')^{-1} \mathbf{j}_T' \\ &= \sigma_\varepsilon^2 \mathbf{I}_T + T\sigma_\alpha^2 \mathbf{P}_T = (\sigma_\varepsilon^2 + T\sigma_\alpha^2) \underbrace{(\mathbf{P}_T + \eta \mathbf{Q}_T)}_{\mathbf{S}_T} \end{aligned} \tag{3.50}$$

where $\mathbf{P}_T \equiv \mathbf{I}_T - \mathbf{Q}_T = \mathbf{j}_T (\mathbf{j}_T \mathbf{j}_T')^{-1} \mathbf{j}_T'$ and $\eta \equiv (\sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + T\sigma_\alpha^2)$. It can be shown that $\mathbf{S}_T^{-1/2} = (1-\lambda)^{-1}[\mathbf{I}_T - \lambda \mathbf{P}_T]$ where $\lambda = (1 - \sqrt{\eta})$ so

$$\mathbf{\Omega}_{\mathsf{RE}}^{-1/2} = (\sigma_\varepsilon^2 + T\sigma_\alpha^2)^{-1/2}(1-\lambda)^{-1}[\mathbf{I}_T - \lambda \mathbf{P}_T] = (1/\sigma_\varepsilon^2)[\mathbf{I}_T - \lambda \mathbf{P}_T]$$

Application of standard GLS gives the transformed equation $\check{\mathbf{y}}_i = \check{\mathbf{X}}_i \boldsymbol{\beta} + \check{\mathbf{u}}_i$ where $\check{\mathbf{q}}_i = ([\mathbf{I}_T - \lambda \mathbf{P}_T]\mathbf{q}_i)$. Then the tth element of $\check{\mathbf{y}}_i$ can be written

$$y_{it} - \lambda \bar{y}_i = (\mathbf{x}_{it} - \lambda \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (u_{it} - \lambda \bar{u}_i). \tag{3.51}$$

i.e., as the FE estimator with λ as an additional quasi-time demeaning parameter.

Summing up: Random and Fixed effects

Random Effects:

- The RE model assumes that the fixed effects are **iid** and **uncorrelated** with the regressors.
- This allows for estimation of the **conditional mean** $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$ and the marginal effect of both **time-constant** and **time-varying** regressors.
- Errors are **serially correlated** and therefore the RE model is usually estimated using **FGLS** or with a robust covariance matrix.

Fixed Effects:

- The FE model is robust to **time-invariant correlation** between regressors and the fixed effect.
- The **conditional mean** $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$ and **marginal effects** of **time-invariant** regressors are not possible to estimate.
- Serially correlated errors can be handled by a **robust covariance matrix** or **FGLS**.

Summing up: Random and Fixed effects

In general, the FE model is preferred since the **key RE assumption** is viewed as **untenable** in many econometric applications.

A classical **Hausman-test** and more flexible variants test the null hypothesis that individual-specific effects are uncorrelated with regressors. Rejection implies one should use the FE model or improve the RE model.

But if hypothesis is not rejected any doubts that the RE assumption holds will persist.

Panel data models

Example 3.7 (Your own research)

In case panel data is available for your research question: which model will use choose?

Outline

3. Panel data models I
    - Motivation
    - Basic panel data concepts and models
    - Basic panel data estimators
    - Application: Comparison of POLS, RE, FE

Example 3.8 (The Effect of Part-time Work on Wages)

OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 77, 4 (2015) 0305–9049
doi: 10.1111/obes.12073

Part-Time Work, Fixed-Term Contracts, and the Returns to Experience*

DANIEL FERNÁNDEZ-KRANZ†, MARIE PAUL‡ and NÚRIA RODRÍGUEZ-PLANAS§,¶

†IE Business School, Lotharstr. 65, Duisburg, NRW 47057, Germany (e-mail: marie.paul@uni-due.de)
‡née Waller, University of Duisburg-Essen, Ruhr Graduate School in Economics, Hohenzollernstr. 1-3, Essen, NRW 45128, Germany (e-mail: marie.paul@uni-due.de)
§Queens College – CUNY 300A Powdermaker Hall, 65-30 Kissena Blvd, Queens, NY 11367, USA (e-mail: nrodriguezplanas@gmail.com)
¶IZA Schaumburg-Lippe-Str, 5-9, Bonn 53113, Germany (e-mail: rodriguez-planas@iza.org)

Abstract

Using data from Spanish Social Security records, we investigate the returns to experience for female workers in different flexible work arrangements. Our model consists of four random-effects equations simultaneously estimated using Markov Chain Monte Carlo techniques. We find a large negative wage effect of working part-time (PT), which differs by motherhood status and contract type. We also find that working PT involves lower returns to experience than standard full-time employment and thus a substantial negative wage effect accumulates over time for those employed PT. Finally, our simulations reveal that working PT also raises the probability of working under a fixed-term contract.

The Effect of Part-time Work on Wages

Source: Example follows the sensitivity analysis in Fernandez-Kranz, Paul, Rodriguez-Planas (*Oxford Bulletin of Economics and Statistics*, 2015).

- The panel data consists of about 15.000 women in Spain observed for 36 quarters.
- Regress log hourly wage on a part-time dummy and controls.
- The main analysis in the paper uses a three equation model endogenizing part-time and fixed-term contract status estimated by MCMC methods and focusses on the employment histories and the interplay of part-time and fixed-term contracts and experience in those.
- In a sensitivity analysis several standard estimators are compared:

	FE	RE	POLS
parttime	-0,061 (0,003)	-0,063 (0,003)	-0,171 (0,008)

- Advantage FE in the application? Disadvantage FE?
- What is surprising about the coefficients? Do you conclude from this that the RE assumption is full-filled in the example? Also consider POLS for your answer.

RE-Estimator as a *Quasi-time-demeaning* Estimator

Remember the derivation in 3.6:

$$\hat{\beta}_{RE} = (\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}'\breve{\mathbf{x}}_{it})^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}'\breve{y}_{it})$$

where $\breve{\mathbf{x}}_{it} = \mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i$, $\breve{y}_{it} = y_{it} - \hat{\lambda}\bar{y}_i$

and $\hat{\lambda} = 1 - \sqrt{(\frac{1}{1+T(\hat{\sigma}_{\alpha}^2/\hat{\sigma}_u^2)}}$.

$\Rightarrow$ Instead of substracting the mean over time from all variables (*time-demeaning*), the RE-estimator substracts only part of the mean (*quasi-time demeaning*).
$\Rightarrow$ Set $\hat{\lambda} = 0$ to obtain the POLS-estimator and $\hat{\lambda} = 1$ to obtain the FE-estimator.

Relation RE and FE

- If $\hat{\lambda}$ is close to one, RE- and FE-estimates are close.
- $\hat{\lambda} \to 1$ if $T \to \infty$ or if $(\hat{\sigma}_\alpha^2/\hat{\sigma}_u^2) \to \infty$. Thus, the estimates are close if T is large or if the estimated variance of α_i is large relative to the estimated variance of u_{it}, thus when the unobserved effect is relatively important.
- In the example a large $\hat{\sigma}_\alpha^2/\hat{\sigma}_u^2$ means that the unexplained part of wages differs strongly between women, so the unobserved effect is relatively important and there is relatively few variation over time.
- $\hat{\sigma}_\alpha^2/(\hat{\sigma}_u^2 + \hat{\sigma}_\alpha^2) = 0.82$, so $\hat{\sigma}_\alpha^2/\hat{\sigma}_u^2 = 4.56$ und $\hat{\lambda} = 0.92$. This explains why FE- and RE results are so close.
- This also implies that the precision of the RE estimator gets close to that of the FE estimator and the coefficients of time constant variables are difficult to estimate.

Relation RE and POLS

- $\hat{\lambda}$ is small if the time constant unobserved effect is realtively unimportant, so if $\hat{\sigma}_u^2$ large relative to $\hat{\sigma}_\alpha^2$ or if T is small. In that case RE and POLS results are similar.
- POLS and RE assumptions both do not allow for correlation between α_i and u_{it}. But in the example we have seen that because of the specific transformation used by the RE estimator its inconsistency can be small relative to POLS, this is if $\hat{\lambda}$ is large.
- The intuition is that POLS leaves α_i completely in the error term and RE only partly.