Lecture 5: Linear Panel Data Model, I

January 28, 2025

Unobserved Effect

- Start with the balanced panel case, and assume random sampling across i (the cross section dimension), with fixed time periods T. So $\{(\mathbf{x}_{it}, Y_{it}) : t = 1, \ldots, T, c_i\}$ where c_i is the unobserved effect drawn along with the observed data.

- The unbalanced case is trickier because we must know why we are missing some time periods for some units. We consider this much later under missing data/sample selection issues.

- For a random draw i from the population, the basic model is

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}, \quad t = 1, \ldots, T,$$

where $\{u_{it} : t = 1, \ldots, T\}$ are the *idiosyncratic errors*. The *composite error* at time t is

$$v_{it} = c_i + u_{it}$$

Assumptions of Unobserved Effect

- As mentioned earlier, we assume a balanced panel and all asymptotic analysis – implicit or explicit – is with fixed T and $N \to \infty$, where N is the size of the cross section.

- The basic *unobserved effects model* is

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}, \quad t = 1, \ldots, T,$$

where $\mathbf{x}_{it}$ is $1 \times K$ and so β is $K \times 1$. In addition to unobserved effect and unobserved heterogeneity, c_i is sometimes called a *latent effect* or an individual effect, firm effect, school effect, and so on.

Assumptions of Unobserved Effect

- In modern applications, "random effect" essentially means

$$\text{Cov}(\mathbf{x}_{it}, c_i) = 0, \quad t = 1, \ldots, T,$$

  although we often will strengthen this.

- The term "fixed effect" means that no restrictions are placed on the relationship between c_i and $\{\mathbf{x}_{it}\}$.

- Recently, "correlated random effects" is used to denote situations where we model the relationship between c_i and $\{\mathbf{x}_{it}\}$, and it is especially useful for nonlinear models (but also for linear models, as we will see).

Assumptions on Explanatory Variables: Strict vs. Sequential Exogeneity

- **Strict Exogeneity Conditional on the Unobserved Effect:**

$$E(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\beta + c_i,$$

  so that only $\mathbf{x}_{it}$ affects the expected value of y_{it} once c_i is controlled for.

- This is weaker than if we did not condition on c_i. Assuming the condition holds conditional on c_i,

$$E(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = \mathbf{x}_{it}\beta + E(c_i|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}).$$

  So correlation between c_i and $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ would invalidate the assumption without conditioning on c_i.

Assumptions on Explanatory Variables: Strict vs. Sequential Exogeneity

- But strict exogeneity conditional on c_i rules out lagged dependent variables and feedback. Written in terms of the idiosyncratic errors, strict exogeneity is

$$E(u_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = 0,$$

and so $\mathbf{x}_{i,t+h}$ must be uncorrelated with u_{it} for all $h > 0$.

- In addition to ruling out feedback, strict exogeneity assumes we have any distributed lag dynamics correct, too. For example, if $\mathbf{x}_{it} = (\mathbf{z}_{it}, \mathbf{z}_{i,t-1})$, then

$$E(y_{it}|\mathbf{z}_{i1}, \ldots, \mathbf{z}_{it}, \ldots, \mathbf{z}_{iT}, c_i) = E(y_{it}|\mathbf{z}_{it}, \mathbf{z}_{i,t-1}, c_i).$$

Assumptions on Explanatory Variables: Strict vs. Sequential Exogeneity

- A more reasonable assumption that we will use later is

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i1}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\beta + c_i,$$

  which is sequential exogeneity conditional on the unobserved effect.

- Sequential exogeneity assumes correct distributed lag dynamics but is silent on feedback.

Alternative Estimator

- Pooled OLS
- "Random" Effect
- "Fixed" Effect – First-Differencing

Estimation: Pooled OLS

- We already covered this. Now, we just recognize that the equation is

$$y_{it} = \mathbf{x}_{it}\beta + v_{it}$$

$$v_{it} = c_i + u_{it}$$

- Consistency (fixed T, $N \to \infty$) of the POLS estimator is ensured by

$$E(\mathbf{x}_{it}'c_i) = 0$$
$$E(\mathbf{x}_{it}'u_{it}) = 0, \quad t = 1, \ldots, T.$$

- Let $\hat{v}_{it} = y_{it} - X_{it}\hat{\beta}_{\text{POLS}}$ be the POLS residuals. Then

$$A\hat{var}(\hat{\beta}_{\text{POLS}}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\hat{v}_{it}'\hat{v}_{ir}\mathbf{x}_{it}'\mathbf{x}_{ir}\right)$$
$$\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}$$

  or sometimes with an adjustment, such as multiply by $N/(N-1)$

- In Stata: **reg y x1 x2 ... xK, cluster(id)**

Random Effects Estimation

- State assumptions in conditional mean terms so that second moment derivations are easier.

Assumption 1 (RE.1)

1. $E(u_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = 0, \quad t = 1, \ldots, T$
2. $E(c_i|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) = E(c_i)$

- Assume $\mathbf{x}_{it}$ includes (at least) unity, and probably time dummies in addition. Then $E(c_i) = 0$ is without loss of generality.

- Write the equation in system form (for all time periods) as

$$Y_i = \mathbf{X}_i\beta + v_i = \mathbf{X}_i\beta + c_i\mathbf{j}_T + \mathbf{u}_i$$

where $\mathbf{j}_T' = (1, 1, \ldots, 1)$.

- Define

$$\mathbf{\Omega}_{T \times T} = E(\mathbf{v}_i\mathbf{v}_i') = \text{Var}(\mathbf{v}_i)$$

Assumption 2 (RE.2)

$\mathbf{\Omega}$ is nonsingular and $rank\ E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$.

Random Effects Estimation

- RE imposes a special structure on Ω (which could be wrong!). Under RE.1(1), c_i and u_{it} are uncorrelated. Assume further that

$$\text{Var}(u_{it}) = \sigma_u^2, \qquad\qquad t = 1, \ldots, T$$
$$\text{Cov}(u_{it}, u_{is}) = 0, \qquad\qquad t \neq s$$

Then

$$\text{Var}(v_{it}) = \text{Var}(c_i + u_{it}) = \text{Var}(c_i) + \text{Var}(u_{it})$$
$$\sigma_v^2 = \sigma_c^2 + \sigma_u^2$$

11

Random Effect Estimation

- Further, for $t \neq s$,

$$\begin{aligned}
\text{Cov}(v_{it}, v_{is}) &= \text{Cov}(c_i + u_{it}, c_i + u_{is}) \\
&= \text{Var}(c_i) + \text{Cov}(c_i, u_{is}) + \text{Cov}(u_{it}, c_i) + \text{Cov}(u_{it}, u_{is}) \\
&= \sigma_c^2
\end{aligned}$$

$\boldsymbol{\Omega}$ is defined as

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \cdots & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{pmatrix},$$

so the $T \times T$ matrix depends on only two parameters, σ_c^2 and σ_u^2 or, more directly, σ_v^2 and σ_c^2.

Random Effect Estimation

- Feasible GLS requires estimating $\boldsymbol{\Omega}$, that is, the two parameters.

- Actually, it would be enough to know $\rho = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_u^2}$, the fraction of the total variance accounted for by c_i. Notice that $\rho = \text{Corr}(v_{it}, v_{is})$ for all $t \neq s$.

- We can also write $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = \sigma_v^2 \begin{pmatrix} 1 & \cdots & \rho & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix},$$

which shows we only need to estimate ρ to proceed with FGLS.

Random Effect Estimation

- We can use pooled OLS to get the residuals, $\hat{v}_{it}$, across all i and t. Then a consistent estimator of σ_v^2 (not generally unbiased), as N gets large for fixed T, is

$$\hat{\sigma}_v^2 = (NT - K)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2 = \frac{SSR}{(NT - K)},$$

- For σ_c^2, note that

$$\sigma_c^2 = [T(T-1)/2]^{-1} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E(v_{it} v_{is}).$$

So a consistent "estimator" would be

$$\hat{\sigma}_c^2 = [NT(T-1)/2]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} v_{it} v_{is}.$$

Random Effect Estimation

- Now we can use

$$
\hat{\boldsymbol{\Omega}} = \begin{pmatrix} \hat{\sigma}_v^2 & \cdots & \hat{\sigma}_c^2 & \hat{\sigma}_c^2 \\ \hat{\sigma}_c^2 & \hat{\sigma}_v^2 & \cdots & \hat{\sigma}_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_c^2 & \cdots & \hat{\sigma}_c^2 & \hat{\sigma}_v^2 \end{pmatrix} \quad \text{or} \quad \hat{\boldsymbol{\Lambda}} = \begin{pmatrix} 1 & \cdots & \hat{\rho} & \hat{\rho} \\ \hat{\rho} & 1 & \cdots & \hat{\rho} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho} & \cdots & \hat{\rho} & 1 \end{pmatrix},
$$

  where $\hat{\rho} = \hat{\sigma}_c^2 / \hat{\sigma}_v^2$ in FGLS.

- The random effects estimator is given by:

$$
\hat{\beta}_{RE} = \left(\sum_{i=1}^{N} \mathbf{X_i}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X_i} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X_i}' \hat{\boldsymbol{\Omega}}^{-1} y_i \right)
$$

- Under assumptions RE.1, RE.2, and RE.3, the asymptotic variance of the random effects estimator:

$$
\hat{Avar}(\hat{\beta}_{RE}) = \left(\sum_{i=1}^{N} \mathbf{X_i}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X_i} \right)^{-1}
$$

  is a valid estimator.

Random Effect Estimation

- Fully robust inference is available for RE, and there are good reasons for doing so:

  1. $\mathbf{\Omega}$ may not have the special (and restrictive, especially for large T) RE structure, that is, $E(\mathbf{v_i v_i})$ need not have the RE form. Serial correlation or changing variances in $\{u_{it} : t = 1, \ldots, T\}$ invalidate the RE structure.
  2. The system homoskedasticity requirement, $E(\mathbf{v_i v_i}|\mathbf{X_i}) = E(\mathbf{v_i v_i})$, might not hold.

- A fully robust estimator is

$$A\hat{var}(\hat{\beta}_{RE}) = \left(\sum_{i=1}^{N} \mathbf{X_i}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X_i}\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X_i}'\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{v_i}}\hat{\mathbf{v_i}}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X_i}\right)$$
$$\cdot \left(\sum_{i=1}^{N} \mathbf{X_i}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X_i}\right)^{-1}$$

where $\hat{\mathbf{v_i}} = y_i - \mathbf{X_i}\hat{\beta}_{RE}$ is the vector of RE (FGLS) residuals.

Random Effect Estimation

In Stata, fully robust inference uses the "cluster" option; for the "usual" variance matrix estimator, drop this option:

xtreg y x1 x2 ... xK, re cluster(id)

Fixed Effects Estimation

- Unlike POLS and RE, fixed effects estimation removes c_i to form an estimating equation.

- Average the original equation,

$$y_{it} = \mathbf{x_{it}}\beta + c_i + u_{it}, \quad t = 1, \ldots, T,$$

across t to get a cross-sectional equation:

$$\bar{y}_i = \mathbf{\bar{x}_i}\beta + c_i + \bar{u}_i,$$

where the overbar indicates time averages:

$$\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}, \quad \mathbf{\bar{x}_i} = T^{-1}\sum_{t=1}^{T} \mathbf{x_{it}}, \quad \bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}.$$

- The equation $\bar{y}_i = \bar{X}_i\beta + c_i + \bar{u}_i$ is often called the *between* equation because it relies on variation in the data between cross section observations. The *between estimator* is the OLS estimator from the cross section regression

$$\bar{y}_i \text{ on } \mathbf{\bar{x}_i}, \quad i = 1, \ldots, N.$$

- In practice, an intercept is included to account for nonzero $E(c_i)$.

Fixed Effects Estimation

- The between estimator is inconsistent unless:

$$\text{Cov}(\bar{\mathbf{x}}_\mathbf{i}, c_i) = 0, \quad \text{Cov}(\bar{\mathbf{x}}_\mathbf{i}, \bar{u}_i) = 0.$$

- Instead, subtract off the time-averaged equation from the original equation to eliminate c_i:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta + u_{it} - \bar{u}_i, \quad t = 1, \ldots, T$$

or

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\beta + \ddot{u}_{it}, \quad t = 1, \ldots, T$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and so on.

- We call this the *time demeaned equation*, and the transformation is time demeaning, fixed effects, or within (time variation within each i is used).

- Key is that c_i is gone from the time demeaned equation. So, we can use pooled OLS:

$$\ddot{y}_{it} \text{ on } \ddot{x}_{it}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N.$$

Fixed Effects Estimation

- Key is that c_i is gone from the time demeaned equation. So, we can use pooled OLS:

$$\ddot{y}_{it} \text{ on } \ddot{\mathbf{x}}_{it}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N.$$

This is the fixed effects (FE) estimator or the within estimator.

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{y}_{it} \right)$$

$$= \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' y_{it} \right),$$

because

$$\sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(y_{it} - \bar{y}_i) = \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}})' y_{it}.$$

Fixed Effects Estimation

Assumption 3 (FE.1)

Same as RE.1(1), that is,

$$E(u_{it}|\mathbf{x}_i, c_i) = 0, \quad t = 1, \ldots, T.$$

- This implies $E(\mathbf{x}'_{is} u_{it}) = 0$, all $s, t = 1, \ldots, T$, and so $E(\ddot{x}'_{it} u_{it}) = 0$, $t = 1, \ldots, T$.

- The rank condition is directly from POLS.2:

Assumption 4 (FE.2)

$$\text{rank}\left(\sum_{t=1}^{T} E(\ddot{x}'_{it} \ddot{x}_{it})\right) = K.$$

- The rank condition rules out elements in $\mathbf{x}_{it}$ that have no time variation for any unit in the population. Such variables get swept away by the within transformation.

Fixed Effects Estimation

- Under FE.1 and FE.2,

$$\hat{\beta}_{FE} \xrightarrow{p} \beta \text{ as } N \to \infty$$

- What parameters can we identify with FE? Suppose we start with

$$y_{it} = \theta_1 + \theta_2 d_{2_t} + \ldots + \theta_T d_{T_t} + \mathbf{z}_i \gamma_1 + d_{2_t} \mathbf{z}_i \gamma_2 + \ldots + d_{T_t} \mathbf{z}_i \gamma_T$$
$$+ W_{it}\delta + c_i + u_{it}$$

- Using FE, we cannot estimate θ_1 or γ_1, but all other parameters are generally identified.

- FE allows c_i to be arbitrarily correlated with $(\mathbf{z}_i, \mathbf{w}_{it})$, and so we cannot distinguish $\theta_1 + \mathbf{z}_i \gamma_1$ from c_i.

- We can estimate $\theta_2, \ldots, \theta_T$ and $\gamma_2, \ldots, \gamma_T$. So we can estimate whether the effect of the time constant variables has changed over time. We cannot estimate the effect in any period t because it is γ_1 for $t = 1$ and $\gamma_1 + \gamma_t$ for $t = 2, \ldots, T$.

Fixed Effects Estimation

- We can obtain a variance matrix estimator valid under Assumptions FE.1 and FE.2.

- Define the FE residuals as

$$\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{x}'_{it}\hat{\beta}_{FE}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N$$

- These are "estimates" of the $\ddot{u}_{it}$, not the u_{it}. This has implications for estimating the error variance, σ_u^2.

- Without additional assumptions, use the "cluster-robust" matrix

$$A\hat{var}(\hat{\beta}_{FE}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}'_{it}\ddot{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\ddot{u}}_{it}\hat{\ddot{u}}_{ir}\ddot{x}'_{it}\ddot{x}_{ir}\right) \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}'_{it}\ddot{x}_{it}\right)^{-1}$$

- In Stata, again use the "cluster" option:
  xtreg y x1 x2 ... xK, fe cluster(id)

Fixed Effects Estimation

Assumption 5 (FE.3)

Same as RE.3(1), that is,

$$E(\mathbf{u_i u_i}'|\mathbf{x_i}, c_i) = \sigma_u^2 \mathbf{I}_T$$

- So under FE.1, FE.2, and FE.3:
$$Avar[\sqrt{N}(\hat{\beta}_{FE} - \beta)] = \sigma_u^2 [E(\mathbf{\ddot{X}}_i' \mathbf{\ddot{X}}_i)]^{-1}$$

- Estimating σ_u^2 requires some care since we effectively observe $\ddot{u}_{it}$, not u_{it}.

- Under the constant variance and no serial correlation assumptions on $\{u_{it}\}$,
$$\text{Var}(\ddot{u}_{it}) = \text{Var}(u_{it} - \bar{u}_i) = \sigma_u^2 + \sigma_u^2/T - 2\text{Cov}(u_{it}, \bar{u}_i)$$
$$= \sigma_u^2 + \sigma_u^2/T - 2\sigma_u^2/T = \sigma_u^2(1 - 1/T)$$

- Hence,
$$\hat{\sigma}_u^2 = [N(T-1) - K]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\ddot{u}}_{it}^2 = \text{SSR}/[N(T-1) - K]$$

24

Fixed Effects Estimation

- The FE estimator $\hat{\beta}_{FE}$ can also be obtained by running a long regression on the original data, and including dummy variables for each cross-section unit:

$$y_{it} = \alpha + d_{1i}, d_{2i}, \ldots, d_{Ni}, \mathbf{x}_{it}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N,$$

often called the *dummy variable regression*. The statistics are properly computed because the inclusion of the N dummy variables.

- Only danger: treating the c_i as parameters to estimate, while sensible with "large" T, can lead to trouble later with nonlinear models. Here, we get a consistent estimator of β for fixed T.

- Sometimes we want to estimate the c_i using the T time periods. Do not have to run the dummy variable regression:

$$\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_{FE}, \quad i = 1, \ldots, N.$$

- With small T, this is not a good "estimate" of c_i, but it is unbiased. We can estimate features of the distribution of c_i well.

Fixed Effects Estimation

- We can estimate other features of the distribution, too, although some "obvious" estimators are inconsistent. For example, we might try to estimate σ_c^2 using the sample variance of $\{\hat{c}_i : i = 1, \ldots, N\}$:

$$\tilde{\sigma}_c^2 = (N-1)^{-1} \sum_{i=1}^{N} (\hat{c}_i - \hat{\mu}_c)^2.$$

- But under FE.1 to FE.3 it can be shown that

$$\text{plim}(\tilde{\sigma}_c^2) = \sigma_c^2 + \text{Var}(\bar{u}_i) = \sigma_c^2 + \sigma_u^2/T.$$

- We can adjust for the "bias" using the estimate $\hat{\sigma}_u^2$:

$$\hat{\sigma}_c^2 = \tilde{\sigma}_c^2 - \hat{\sigma}_u^2/T = (N-1)^{-1} \sum_{i=1}^{N} (\hat{c}_i - \hat{\mu}_c)^2 - \hat{\sigma}_u^2/T$$

is consistent for σ_c^2 for any T as $N \to \infty$.

Practical Hints in Applyting Fixed Effects

- Possible confusion concerning the term "fixed effects." Suppose i is a firm. Then the phrase "firm fixed effect" corresponds to allowing c_i in the model to be correlated with the covariates. If c_i is called a firm "random effect," then it is being assumed to be uncorrelated with $\mathbf{x}_{it}$.

- Suppose that we cannot, or do not want to, use FE estimation. This might occur because the key variable at the firm level is constant across time for all firms – and so the FE transformation sweeps it away – or there is little time variation within firm in the key variable, leading to large standard errors.

Practical Hints in Applying Fixed Effects

- Instead, we might use a random effects analysis at the firm level but include industry dummy variables to account for systematic differences across industries. So, we include in $\mathbf{x}_{it}$ a set of industry dummy variables while also allowing a firm effect c_i in a "random effects" analysis.

- If there are many firms per industry, the industry "fixed effects" – the coefficients on the industry dummies – can be precisely estimated. So the industry "fixed effects" are really parameters to estimate whereas the c_i are not.

- Generally, including dummies for more aggregated levels and then applying RE is common when the covariates of interest vary in the cross-section but not (much) over time.

- Keep in mind that an RE analysis at the firm level with industry dummies need not be entirely convincing: the key elements of $\mathbf{x}_{it}$ might be correlated with unobserved firm features that are not adequately captured by industry differences.