

5 Panel Data Models

5.1 Panel Data Setup

A general representation of a linear panel data model is given by:

$$y_{it} = \alpha_i + \beta'_{1i}x_{1it} + \beta'_{2i}x_{2i} + \beta'_{3i}x_{3t} + \varepsilon_{it} \quad \text{for } i = 1, \dots, N \text{ and } t = 1, \dots, T$$

α_i : individual specific effect, E.G. unobserved institutional effects in a cross country growth model
 x_{1it} : $k_1 \times 1$ vector of measures varying over T and N, e.g. individual income
 x_{2i} : $k_2 \times 1$ vector of measures varying over N, e.g. individual education
 x_{3t} : $k_3 \times 1$ vector of measures varying over T, e.g. interest rates

Note that x_{2i} is distinct from α_i , the former represents *observed heterogeneity* across individuals, while the latter attempts to partially control for *unobserved heterogeneity* across individuals.

We can impose that the parameters β are identical across individuals to get the *homogenous slope panel model*, which we focus on for the rest of the lecture.

$$y_{it} = \alpha_i + \beta'x_{it} + \varepsilon_{it}, \quad \text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$$

$$x_{it} = \begin{bmatrix} x_{1it} \\ x_{2i} \\ x_{3t} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Or in matrix form:

$$y_i = \alpha_i \mathbf{1}_T + X_i \beta + \varepsilon_i$$

Definition 5.1.1: Strict exogeneity

$$\mathbb{E}[\varepsilon_{it} | x_{i1}, x_{i2}, \dots, x_{iT}, \alpha_i] = \mathbb{E}[\varepsilon_{it}] = 0$$

This implies $\text{Cov}(x_{it}, \varepsilon_{is}) = 0 \quad \forall i, t, s$. Intuitively: every error is uncorrelated with every regressor at every time period.

Example (A violation of strict exogeneity). Consider a model where ε_{it} is correlated with future values of the regressors, e.g.

$$x_{i,t+1} = \varphi \varepsilon_{it} + e_{it+1}$$

A production function in which labour demand in period $t + 1$ responds to unobserved productivity shocks in period t , represented by ε_{it} , would be an example of this.

Definition 5.1.2: Weak exogeneity

$$\mathbb{E}[\varepsilon_{it} | x_{i1}, x_{i2}, \dots, x_{it}, \alpha_i] = 0$$

This implies $\text{Cov}(x_{it}, \varepsilon_{is}) = 0 \quad \forall s \geq t$. Intuitively: every error is uncorrelated with every regressor at or before the current time period.

Example (Programme Evaluation and Wages). Many studies have considered the impact of job training programmes on wages. A standard specification is:

$$\log(wage_{it}) = \alpha_i + x'_{it}\beta + \delta prog_{it} + \varepsilon_{it}$$

Consider a 2 period model where at $t=1$ nobody is treated and at $t=2$ a group of individuals enter the program, and wages are observed for both treated and untreated individuals. Do we have strict exogeneity?

Is there correlation between $prog_{it}$ and ε_{it} conditional on α_i ? We might think not, an individual cannot react to wage shock and enlist in program in the same period.

Is there correlation between $prog_{it+1}$ and ε_{it} ? This might be non-zero if future participation is based on current shocks to wages.

Note:-

A general dynamic panel model might be written as:

$$y_{it} = \tau y_{i,t-1} + \beta' x_{it} + \alpha_i + \mu_t + \varepsilon_{it}$$

where μ_t is a time specific effect. We have an identification problem here, since $y_{i,t}$ depends on the fixed term α_i , so does $y_{i,t-1}$ which implies non-zero correlation between $y_{i,t-1}$ and the error $\alpha_i + \mu_t + \varepsilon_{it}$. This is true even if ε_{it} is serially uncorrelated (but it is clear to see that the correlation grows with serial correlation in ε_{it}). To estimate this model we need to use *instrumental variables* or *GMM*.

Define the composite error $v_{it} = \alpha_i + \varepsilon_{it}$. We are thus interested in estimation of β in the model $y_{it} = x'_{it}\beta + v_{it}$. We consider three approaches to estimation under this setup: *pooled OLS*, *fixed effects*, and *random effects*.

5.2 Pooled OLS (POLS)

Definition 5.2.1

The pooled OLS estimator is given by:

$$\hat{\beta}_{POLS} = \left(\sum_{i=1}^N \sum_{t=1}^T x_{it} x'_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T x_{it} y_{it} \right) = \left(\sum_{i=1}^N X'_i X_i \right)^{-1} \left(\sum_{i=1}^N X'_i y_i \right)$$

where $\bar{x} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it}$ and $\bar{y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it}$.

Assumptions:

$$(POLS1a) \quad \mathbb{E}[x'_{it}\alpha_i] = 0 \quad \forall t$$

$$(POLS1b) \quad \mathbb{E}[x'_{it}\varepsilon_{it}] = 0 \quad \forall t$$

$$(POLS2) \quad \text{rank}(\mathbb{E}[x_{it} x'_{it}]) = k \quad \forall i$$

(POLS1) is equivalent to the standard exogeneity assumption, but on the composite error: $\mathbb{E}[x'_{it}v_{it}]$. POLS is consistent under (POLS1) and (POLS2), however the composite errors will be serially correlated on account of the common α_i term. Indeed this correlation does not decrease as $|t-s|$ increases, therefore inference using POLS requires robust standard errors.

5.3 Fixed Effects

In many cases we might expect the unobserved effect α_i to be correlated with the observed x_{it} , in which case POLS is inconsistent. We can instead use fixed effects in this case. Consider the model in matrix form from above:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1_T & 0 & \dots & 0 \\ 0 & 1_T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1_T \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

The *Least Squares Dummy Variable* (LSDV) estimator can be computed by running OLS on the stacked equations, giving N fixed effects. However, as N gets large, the inversion of a matrix of dimension $(N+k) \times (N+k)$ becomes computationally infeasible. The Fixed effects transformation by contrast only requires the inversion of a $k \times k$ matrix.

Definition 5.3.1

The fixed effects estimator is given by:

$$\begin{aligned} \hat{\beta}_{FE} &= \left(\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \right) \\ &= \left(\sum_{i=1}^N (Q_T X_i)' (Q_T X_i) \right)^{-1} \left(\sum_{i=1}^N (Q_T X_i)' (Q_T y_i) \right) \end{aligned}$$

where $Q_T = I_T - \frac{1}{T} \mathbf{1}_T \mathbf{1}_T'$ and $\bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$.

Assumptions:

- (FE1) $\mathbb{E}[\varepsilon_{it} | X_i, \alpha_i] = 0 \quad \forall t$
- (FE2) $\text{rank}(\mathbb{E}[X_i' Q_T X_i]) = k$
- (FE3) $\mathbb{E}[\varepsilon_i \varepsilon_i' | X_i, \alpha_i] = \sigma_\varepsilon^2 I_T$

FE1:

Strict exogeneity conditional on the unobserved effects, however note that $\mathbb{E}[\alpha_i | X_i]$ is allowed to be any function of X_i . The cost of this assumption is that we lose the ability to estimate individual specific effects, as there is no way to distinguish between the time invariant unobserved and observed effects.

FE2:

If x_{it} contains an element that is constant over time for any i , then the corresponding column of $X_i' Q_T X_i$ will be zero, and the rank condition will not hold.

FE3:

Constant conditional variance and zero conditional covariance of the idiosyncratic error.

The matrix Q_T is an idempotent matrix with all off-diagonal elements equal to $-\frac{1}{T}$ and all diagonal elements equal to $1 - \frac{1}{T}$. $Q_T y_i$ simply subtracts individual means from y_i .

The FE for β is consistent (and unbiased) under (FE1) and (FE2) with both $N \rightarrow \infty$ and $T \rightarrow \infty$. Estimates of the fixed effects can be obtained using $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i' \hat{\beta}_{FE}$. The FE estimate of α_i is unbiased and consistent for fixed N and $T \rightarrow \infty$, however we have an incidental parameters problem with increasing N as the number of fixed effects grows with N .

5.4 Random Effects

If we assume $\alpha_i \sim iid(0, \sigma_\alpha^2)$ independent of $\varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$ and uncorrelated with x_{it} , we have a random effects or one-way error-components setup. We effectively put α_i into the error term, under the assumption that it is orthogonal to x_{it} , then accounts for the implied serial correlation with GLS.

Definition 5.4.1

The random effects estimator is a FGLS estimator, given by

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

Assumptions:

$$(RE1a) \quad \mathbb{E}[\varepsilon_{it}|X_i, \alpha_i] = 0 \quad \forall t$$

$$(RE1b) \quad \mathbb{E}[\alpha_i|X_i] = \mathbb{E}[\alpha_i] = 0 \quad \forall t$$

$$(RE2) \quad rank(\mathbb{E}[X_i' \Sigma^{-1} X_i]) = k$$

$$(RE3a) \quad \mathbb{E}[\varepsilon_i \varepsilon_i' | X_i, \alpha_i] = \sigma_\varepsilon^2 I_T$$

$$(RE3b) \quad \mathbb{E}[\alpha_i^2 | X_i] = \sigma_\alpha^2$$

RE1:

(RE1a) is the strict exogeneity assumption (**exactly the same as FE1**), while (RE1b) is the orthogonality assumption, this is implied by the assumption that x_{it} are fixed and $\mathbb{E}[\alpha_i|X_i] = 0$ or that α_i is independent of X_i . The important part of this assumption is $\mathbb{E}[\alpha_i|X_i] = \mathbb{E}[\alpha_i]$, that it equals zero is without loss of generality provided an intercept is included in x_{it} .

Why do we make (RE1), which is much stronger than (POLS1)? Random effects exploits serial correlation in the composite error in a GLS framework, to ensure it's consistent we need some form of strict exogeneity between the composite error and the regressors.

RE2:

For consistency we need the usual rank condition for GLS, we know GLS and FGLS are consistent under RE1 and RE2. However we have not yet exploited the unobserved effects structure of v_{it} .

RE3:

(RE3a) is constant conditional variance and zero conditional covariance of the idiosyncratic error (**exactly the same as FE3**), this is stronger than constant unconditional variance. I.e.: $\mathbb{E}[\varepsilon_{it}^2 | X_i, \alpha_i] = \sigma_\varepsilon^2 \quad \forall t$ and $\mathbb{E}[\varepsilon_{it} \varepsilon_{is} | X_i, \alpha_i] = 0 \quad \forall t \neq s$.

(RE3b) is a homoskedasticity assumption on the unobserved effect. Taken together these imply the unconditional assumptions:

$$\begin{array}{llll} \mathbb{E}[\alpha_i | X_i] = 0 & \mathbb{E}[\alpha_i^2 | X_i] = \sigma_\alpha^2 & \mathbb{E}[\alpha_i \alpha_j | X_i] = 0 & \forall i \neq j \\ \mathbb{E}[\varepsilon_{it} | X_i] = 0 & \mathbb{E}[\varepsilon_{it}^2 | X_i] = \sigma_\varepsilon^2 & \mathbb{E}[\varepsilon_{it} \varepsilon_{js} | X_i] = 0 & \forall i \neq j, t \neq s \\ \mathbb{E}[\alpha_i \varepsilon_{jt} | X_i] = 0 & \forall i, j, t & & \end{array}$$

We can now write the unconditional covariance matrix of v_{it} as:

$$\begin{aligned}
\mathbb{E}[v_i v_i'] &= \mathbb{E}[(\alpha_i + \varepsilon_i)(\alpha_i + \varepsilon_i)'] \\
&= \mathbb{E}[\alpha_i \alpha_i'] + \mathbb{E}[\alpha_i \varepsilon_i'] + \mathbb{E}[\varepsilon_i \alpha_i'] + \mathbb{E}[\varepsilon_i \varepsilon_i'] \\
&= \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_\varepsilon^2 I_T \\
&= \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\varepsilon^2 + \sigma_\alpha^2 \end{bmatrix}
\end{aligned}$$

Rather than depending on $T(T+1)/2$ unrestricted covariances (as in normal GLS), Σ only depends on 2 parameters, regardless of T . The correlation between composite errors does not depend on the size of the difference between t and s , and is given by $\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$.

$\mathbb{E}[v_i v_i']$ is written in terms of single individual, we can write the $NT \times NT$ covariance matrix for all individuals as:

$$\Omega = I_n \otimes \Sigma = \begin{bmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma \end{bmatrix}$$

We can write this as block diagonal due to random sampling in the cross section (i.e. observations i and j are independent). This is convenient as we can invert Ω by focusing on inverting Σ .

Lemma 5.4.1 (Inverse of a block diagonal matrix).

$$\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n \end{bmatrix}^{-1} = \begin{bmatrix} A_1^{-1} & 0 & \dots & 0 \\ 0 & A_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_n^{-1} \end{bmatrix}$$

Observe that by transforming the regression by premultiplying with $\Sigma^{-1/2}$, we can see $\text{Var}(\Sigma^{-1/2} v_i) = I_T$ and $\mathbb{E}[v_i v_j] = 0$ for $i \neq j$.

Lemma 5.4.2 (Matrix inversion lemma). If a and b are nonzero scalars and P and Q are symmetric idempotent matrices such that $PQ = QP = 0$ and $P + Q = I$, then:

$$(aP + bQ)^{-1} = \frac{1}{a}P + \frac{1}{b}Q$$

Theorem 5.4.1. The GLS transformation

$$\Sigma^{-1/2} y_i = \Sigma^{-1/2} \mathbf{1}_T \lambda + \Sigma^{-1/2} X_i \beta + \Sigma^{-1/2} v_i$$

takes the form of a quasi-demeaning transformation, corresponding to an OLS regression of $y_{it}^* \equiv y_{it} - \theta \bar{y}_i$ on $x_{it}^* \equiv x_{it} - \theta \bar{x}_i$ where $\theta = 1 - \frac{\sigma_\alpha^2}{\sqrt{T\sigma_\varepsilon^2 + \sigma_\alpha^2}}$.

Proof. We know:

$$\Sigma = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_\varepsilon^2 I_T$$

Define $P = \frac{\mathbf{1}_T \mathbf{1}_T'}{T}$ and $Q = I_T - \frac{\mathbf{1}_T \mathbf{1}_T'}{T}$. Clearly these are the residual maker and projection matrices onto the space spanned by the constant vector, and thus have all the required properties to apply the lemma.

$$\begin{aligned}\Sigma &= \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_\varepsilon^2 I_T \\ &= T\sigma_\alpha^2 P + \sigma_\varepsilon^2 (P + Q) \\ &= (T\sigma_\alpha^2 + \sigma_\varepsilon^2)P + \sigma_\varepsilon^2 Q \\ \Rightarrow \Sigma^{-1} &= \frac{1}{T\sigma_\alpha^2 + \sigma_\varepsilon^2} P + \frac{1}{\sigma_\varepsilon^2} Q \\ &= \frac{1}{T\sigma_\alpha^2 + \sigma_\varepsilon^2} (P + \psi^{-1}Q) \quad \text{where } \psi = \frac{\sigma_\varepsilon^2}{T\sigma_\alpha^2 + \sigma_\varepsilon^2}\end{aligned}$$

$$\begin{aligned}\text{note that } (P + \psi^{-1/2}Q)^2 &= P + \psi^{-1/2}PQ + \psi^{-1/2}QP + \psi^{-1}Q \\ &= P + \psi^{-1}Q \quad \text{since } PQ = QP = 0\end{aligned}$$

$$\begin{aligned}\text{thus } \Sigma^{-1/2} &= \frac{1}{\sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}} (P + \psi^{-1/2}Q) \\ &= \frac{1}{\sigma_\varepsilon} \psi^{1/2} (\psi^{-1/2}I_T + (1 - \psi^{-1/2})P) \\ &= \frac{1}{\sigma_\varepsilon} (I_T - (1 - \psi^{1/2})P)\end{aligned}$$

Define $\theta = 1 - \frac{\sigma_\alpha^2}{\sqrt{T\sigma_\alpha^2 + \sigma_\varepsilon^2}}$. Note that P is a $T \times T$ matrix with every element equal to $\frac{1}{T}$. We apply the transformation to y_i :

$$\begin{aligned}\Sigma^{-1/2} y_i &= \frac{1}{\sigma_\varepsilon} (I_T - \theta P) y_i \\ &= \frac{1}{\sigma_\varepsilon} \begin{bmatrix} 1 - \frac{\theta}{T} & -\frac{\theta}{T} & \dots & -\frac{\theta}{T} \\ -\frac{\theta}{T} & 1 - \frac{\theta}{T} & \dots & -\frac{\theta}{T} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\theta}{T} & -\frac{\theta}{T} & \dots & 1 - \frac{\theta}{T} \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \\ &= \frac{1}{\sigma_\varepsilon} \begin{bmatrix} y_{i1} - \frac{\theta}{T} \sum_{t=1}^T y_{it} \\ y_{i2} - \frac{\theta}{T} \sum_{t=1}^T y_{it} \\ \vdots \\ y_{iT} - \frac{\theta}{T} \sum_{t=1}^T y_{it} \end{bmatrix} \\ &= \frac{1}{\sigma_\varepsilon} \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ \vdots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix}\end{aligned}$$

□

Note that the GLS estimator itself is not implementable since σ_α^2 and σ_ε^2 are unknown. We instead estimate $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\varepsilon^2$ and perform FGLS. Note that this estimator may be inefficient for a number

of reasons, for example ε_{it} may be heteroskedastic over time or serially correlated. To combat this (and for fixed T and large N), a robust covariance matrix should be used.

Explanation. *The random effects (equicorrelated) panel model may be viewed as a Pooled model, with the random effect subsumed within the error term*

In the random effects set up the composite error is independent of the regressors, so we could indeed run POLS with the composite error. Random effects performs a GLS transformation to account for the specific form of correlation in the composite error, and is subsequently efficient. \square

Explanation. *How might we generalise the FGLS estimator which follows from RE1-RE3?*

If the idiosyncratic errors are generally heteroskedastic and serially correlated across t , a more general estimator of Ω could be used:

$$\hat{\Omega} = N^{-1} \sum_{i=1}^N \hat{v}_i \hat{v}_i'$$

where \hat{v}_i is the POLS residual. We would not use this normally since it has $T(T+1)/2$ estimated elements (v.s. 2 for RE), however if N is sufficiently large we can use this to perform FGLS instead. \square

Explanation. *Inference should be based on panel-robust standard errors (SE) that permit errors to be correlated over time for a given individual and allow variances and covariances to differ across i .*

Default SE assumes independence of model errors over t for a given i ; this overestimates the benefit of an additional T resulting in downward biased SEs. In reality there is correlation across time for a given i so an extra T is less informative than if it was independent.

Ignoring heteroskedasticity in errors also leads to biased SEs, though this could be in either direction. \square

5.4.1 RE as a nested model

We can see how the random effects estimator nests both the POLS and FE estimators by examining the limits of θ :

POLS: $\theta = 0$

When $\psi = 1$ (because $\sigma_\alpha = 0$) there is no error covariance across observations for a given i , so GLS reduces to OLS. This is because the fixed effects are absorbed into the constant term (since they are the same for all individuals), and the error term is just the idiosyncratic error.

FE: $\theta = 1$

When $\psi = 0$ (because σ_α or $T \rightarrow \infty$) GLS converges to FE as we give more weight to within individual sample means, and quasi-demeaning becomes full demeaning.

Case 1 $\sigma_\alpha \rightarrow \infty$: The effect of time-constant explanatory variables becomes harder to estimate due to noise in the unobserved component.

Case 2 $T \rightarrow \infty$: The unobserved α_i becomes observable, we know our estimator of $[\alpha, \beta]$ is consistent in T or n, thus:

$$y_{it} - x'_{it}\beta = \alpha_i + \varepsilon_{it}$$

becomes observable. The individual means give:

$$\bar{y}_i - \bar{x}'_i\beta = \alpha_i + \bar{\varepsilon}_i$$

and by LLN $\bar{\varepsilon}_i \rightarrow 0$, revealing α_i .

5.5 Comparing estimators

A useful way of comparing estimators is by examining how they use variation within and between individuals. We can decompose the total variation of a series as¹:

$$\begin{aligned}
 T_{yy} &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y})^2 \\
 &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i + \bar{y}_i - \bar{y})^2 \\
 &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i)^2 + \sum_{i=1}^N \sum_{t=1}^T (\bar{y}_i - \bar{y})^2 + 2 \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i)(\bar{y}_i - \bar{y}) \\
 &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i)^2 + \sum_{i=1}^N \sum_{t=1}^T (\bar{y}_i - \bar{y})^2 \\
 &\quad \text{within} \qquad \qquad \text{between}
 \end{aligned}$$

POLS: Minimises the total sum of squares, weighting within and between variation equally.

FE: Disregards between variation, only uses variation within individuals.

RE: Uses both within and between variation, but never weighting between more than within.

We can see a bias-variance tradeoff here, we can completely drop between variation (resulting in a higher variance estimate) to get unbiasedness through dropping α_i as in fixed effects, or we can use both within and between variation to get a lower variance estimate, but potentially with a bias from the unobserved effects. We can think of RE as optimally picking these weights under certain assumptions on the structure of the unobserved effects.

The fundamental differences come from how we treat the unobserved effects, as shown below.

	Constant $\alpha_i = \alpha$	Random $\alpha_i \sim (\alpha, \sigma_\alpha^2)$	Fixed $\mathbb{E}[\alpha_i X_i] \neq 0$
POLS	BLUE Consistent	Unbiased Consistent Inefficient	Biased Inconsistent
FE	Unbiased Consistent Inefficient	Unbiased Consistent Inefficient	BLUE Consistent
RE	Unbiased Consistent Efficient(?)	BLUE Consistent	Biased Inconsistent

5.5.1 Hausman test

We can test whether the random effects specification is supported by the data. The basis for this is that RE is only consistent under the assumption that α_i is uncorrelated with the regressors,

¹the final equality follows from $\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i)(\bar{y}_i - \bar{y}) = \sum_{i=1}^N \left((\bar{y}_i - \bar{y}) \sum_{t=1}^T (y_{it} - \bar{y}_i) \right) = \sum_{i=1}^N (\bar{y}_i - \bar{y})(T\bar{y}_i - T\bar{y}) = 0$

whereas FE is always consistent regardless of the conditional distribution of α_i . We thus test on the difference between the RE and FE estimators, which should be zero if RE is correctly specified. We want to test: $H_0 : \mathbb{E}[\alpha_i|X_i] = 0$ against $H_1 : \mathbb{E}[\alpha_i|X_i] \neq 0$. We can summarise our estimators:

Estimator	Under H_0	Under H_1
POLS	Consistent	Inconsistent
FE	Consistent	Consistent
RE	Efficient	Inconsistent
Between	Consistent	Inconsistent

Define :

$$\hat{\kappa} = \hat{\beta}_{RE_s} - \hat{\beta}_{FE_s}$$

Where $\hat{\beta}_{RE_s}$ is the subcomponent of $\hat{\beta}_{RE}$ relating to time-varying regressors. Under H_0 , $\hat{\kappa} \xrightarrow{p} 0$ and $Cov(\hat{\kappa}, \hat{\beta}_{FE_s}) = 0$ (???), thus $Var(\hat{\kappa}) = Var(\hat{\beta}_{FE_s}) - Var(\hat{\beta}_{RE_s})$ (???). The Hausman test statistic is then:

$$H = \hat{\kappa}' (Var(\hat{\kappa}))^{-1} \hat{\kappa} \sim \chi^2(k_s)$$

where k_s is the number of time-varying regressors.

Note:-

Problems with the Hausman test

- This form of the Hausman test is invalid if α_i or ε_{it} are not i.i.d.
- There is no guarantee that $Var(\hat{\kappa})$ is positive definite.

A moment based test of the RE model can be used given that the model is over-identified. We know FE is consistent under RE1 (since RE1a = FE1) but we can test whether the extra moment condition from RE1b is satisfied by an overid test.

Example. Suppose $\hat{\beta}_{FE} = 0.168$ and $\hat{\beta}_{RE} = 0.119$ with standard deviations 0.019 and 0.014 respectively, is there evidence of correlation between α_i and X_i ?

$$H = \frac{(0.168 - 0.119)^2}{0.019^2 - 0.014^2} \approx 14 > \chi_{.05}^2(1) = 3.84$$

so the random effects model is rejected. The statistic H seems inflated because we have used default standard errors (which are greatly downward biased). This is a signal that a more general form of the Hausman test should be used.

5.6 Correlated Random Effects

We can extend the random effects model to allow for correlation between the unobserved effects and the regressors. Given random sampling in the cross section, an analyst starts with the following conditional mean assumption:

$$\mathbb{E}[\varepsilon_{it}|x_{i1}, x_{i2}, \dots, x_{iT}, \alpha_i] = 0 \quad \forall t$$

Further we now impose the additional moment condition²:

$$\mathbb{E}[\alpha_i|X_i] = \psi + \bar{x}_i'\delta$$

²we are now overidentified

where $\alpha_i = \psi + \bar{x}'_i \delta + \nu_i$ and $\mathbb{E}[\nu_i|X_i] = 0$. Note that previously we were only using $\alpha_i = \psi + \nu_i$, now we are allowing for a linear relationship between α_i and \bar{x}_i . **Thus when $\delta = 0$, we have the standard random effects model.**

Given this, we can find an expression for $\mathbb{E}[y_{it}|X_i]$:

$$\begin{aligned} y_{it} &= \alpha_i + \beta' x_{it} + \varepsilon_{it} \\ &= \psi + x'_{it} \beta + \bar{x}'_i \delta + \nu_i + \varepsilon_{it} \\ \Rightarrow \mathbb{E}[y_{it}|X_i] &= \psi + x'_{it} \beta + \bar{x}'_i \delta \end{aligned}$$

Assumptions:

$$\mathbb{E}[\nu_i|X_i] = 0 \quad \mathbb{E}[\varepsilon_{it}|X_i] = 0 \quad \forall t$$

This allows us to keep the modelling structure of RE (with α_i randomly distributed), but allows for some correlation between α_i and x_{it} . By adding \bar{x}_i we control for some correlation between α_i and x_{it} , the remainder ν_i is uncorrelated with x_{it} .

Under the assumptions above, the estimating equation is given by:

$$\begin{aligned} y_{it} &= \psi + x'_{it} \beta + \bar{x}'_i \delta + \nu_i + \varepsilon_{it} \\ &= \psi + x'_{it} \beta + \bar{x}'_i \delta + u_{it} \end{aligned}$$

and we can just use POLS to consistently estimate all parameters, including δ . Further we can include time constant variables, for example if we begin with

$$y_{it} = \alpha_i + g'_t \theta + z'_i \zeta + x'_{it} \beta + \varepsilon_{it}$$

we can use the CRE estimating equation

$$y_{it} = g'_t \theta + z'_i \zeta + x'_{it} \beta + \psi + \bar{x}'_i \delta + u_{it}$$

which is algebraically equivalent to a fixed effects transformation! Thus we get the FE estimates of θ and β (the coefficients on the time varying covariates). This equivalence implies an interesting interpretation of the FE estimator: it controls for the average level \bar{x}_i when measuring the partial effects of x_{it} on y_{it} . It would not be necessary to control for this average level of \bar{x}_i , as with RE, if $\delta = 0$. A test of $H_0 : \delta = 0$ is then a test of the RE specification.