Chapter 1. Linear Panel Models and Heterogeneity
Master of Science in Economics - University of Geneva

Christophe Hurlin, Université d’Orléans

Université d’Orléans

January 2010
The outline of the chapter:

1. **Specification tests and analysis of covariance**
2. **Linear unobserved effects panel data models**
3. **Fixed or random methods?**
4. **Fixed-Effects methods: Least Squares dummy variable approach**
5. **Random effects methods**
6. **Heterogeneous panels: random coefficients models**
Section 1.
Specification tests and analysis of covariance
A linear model commonly used to assess the effects of both quantitative and qualitative factors is postulated as

\[
y_{i,t} = \alpha_{i,t} + \beta'_{i,t}x_{i,t} + \varepsilon_{i,t}
\]

where \( \forall i = 1, \ldots, N, \forall t = 1, \ldots, T \) 
\( \alpha_{it} \) and \( \beta = (\beta_{1it}, \beta_{2it}, \ldots, \beta_{Kit}) \) are \( 1 \times 1 \) and \( 1 \times K \) vectors of parameters that vary across \( i \) and \( t \), 
\( x_{it} = (x_{1it}, \ldots, x_{Kit}) \) is a \( 1 \times K \) vector of exogenous variables, 
\( u_{it} \) is the error term.
Three aspects of the estimated regression coefficients can be tested:

1. the homogeneity of regression slope coefficients
2. the homogeneity of regression intercept coefficients.
3. the time stability of parameters (slopes and constants). We will not consider this issue (not specific to panel data models) here.
We assume that parameters are **constant over time**, but can vary across individuals.

\[ y_{i,t} = \alpha_i + \beta_i' x_{i,t} + \varepsilon_{i,t} \]

Three types of restrictions can be imposed on this model. Regression slope coefficients are identical, and intercepts are not (model with individual / unobserved effects).

\[ y_{i,t} = \alpha_i + \beta' x_{i,t} + \varepsilon_{i,t} \]

Regression intercepts are the same, and slope coefficients are not (unusual).

\[ y_{i,t} = \alpha + \beta_i' x_{i,t} + \varepsilon_{i,t} \]

Both slope and intercept coefficients are the same (homogeneous / pooled panel).

\[ y_{i,t} = \alpha + \beta' x_{i,t} + \varepsilon_{i,t} \]
Definition

An **heterogeneous panel data model** is a model in which all parameters (constant and slope coefficients) vary across individuals.

Definition

An **homogeneous panel data model (or pooled model)** is a model in which all parameters (constant and slope coefficients) are common.
Specification Tests
Specification tests and analysis of covariance

Linear unobserved effects panel data models
Random coefficients models

Application: strikes in OECD

Figure: Hsiao (2003)
Speciﬁcation tests and analysis of covariance

Linear unobserved effects panel data models
Random coefﬁcients models

Application: strikes in OECD

Lemma

Under the assumption that the \( \varepsilon_{it} \) are independently normally distributed over \( i \) and \( t \) with mean zero and variance \( \sigma_{\varepsilon}^2 \), \( F \) tests can be used to test the restrictions postulated
First step (homogeneous/pooled assumption)

Let us consider the general model

$$y_{i,t} = \alpha_i + \beta' x_{i,t} + \varepsilon_{i,t}$$

The hypothesis of common intercept and slope can be viewed as a set of \((K+1)(N-1)\) linear restrictions:

$$H^1_0 : \beta_i = \beta \quad \alpha_i = \alpha \quad \forall i \in [1, N]$$

$$H^1_a : \exists (i, j) \in [1, N] \ / \ \beta_i \neq \beta_j \text{ ou } \alpha_i \neq \alpha_j$$
Under the alternative $H_1$, there are $NK$ estimated slope coefficients for the $N$ vectors $\beta_i$ ($K \times 1$) and estimated $N$ constants.

Under $H_1$, the unrestricted residual sum of squares $S_1$ divided by $\sigma^2_\varepsilon$ has a chi-square distribution with $NT - N(K + 1)$ degrees of freedom.
Under the homogeneous assumption $H_0^1$, 

$$H_0^1 : \beta_i^{(K,1)} = \beta \quad \alpha_i^{(K,1)} = \alpha \quad \forall \ i \in [1, N]$$

the F statistic, denoted $F_1$, and defined by:

$$F_1 = \frac{(RSS_{1,c} - RSS_1)}{RSS_1} \times \frac{\left( (N - 1) (K + 1) \right)}{[NT - N (K + 1)]}$$

has a Fischer distribution with $(N - 1) (K + 1)$ and $NT - N (K + 1)$ degrees of freedom. $RSS_1$ denotes the residual sum of squares of the model and $RSS_{1,c}$ the residual sum of squares of the constrained model.
Specification tests and analysis of covariance

Linear unobserved effects panel data models
Random coefficients models

Application: strikes in OECD

Under $H_1$, the residual sum of squares is equal to the sum of the $N$ residual sum of squares associated to the $N$ individual regressions:

$$RSS_1 = \sum_{i=1}^{N} RSS_{1,i} = \sum_{i=1}^{N} \left[ S_{yy,i} - S_{xy,i}S_{xx,i}^{-1}S_{xy,i} \right]$$

with

$$S_{yy,i} = \sum_{t=1}^{T} (y_{i,t} - \bar{y}_i)^2 \quad \text{with} \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t} \quad \text{and} \quad \bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}$$

$$S_{xx,i} = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(x_{i,t} - \bar{x}_i)'$$

$$S_{xy,i} = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(y_{i,t} - \bar{y}_i)'$$
Under $H_0^1$, the model is:

$$y_{i,t} = \alpha + \beta' x_{i,t} + \varepsilon_{i,t}$$

the least-squares regression of the pooled model yields parameter estimates

$$\hat{\beta} = S_{xx}^{-1} S_{xy}$$

$$S_{xx} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x})(x_{i,t} - \bar{x})' \quad \text{with} \quad \bar{x} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{i,t}$$

$$S_{xy} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x})(y_{i,t} - \bar{y})' \quad \text{with} \quad \bar{y} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t}$$
Under $H_0^1$, the overall RSS is defined by

$$SCR_{1,c} = S_{yy} - S_{xy}'S_{xx}^{-1}S_{xy}$$

with

$$S_{yy} = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{i,t} - \bar{y}_i)^2$$

$$S_{xx,i} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(x_{i,t} - \bar{x}_i)'$$

$$S_{xy,i} = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(y_{i,t} - \bar{y}_i)'$$
Second step (individual/unobserved effects)

Let us consider the general model

\[ y_{i,t} = \alpha_i + \beta_i' x_{i,t} + \varepsilon_{i,t} \]

The hypothesis of heterogeneous intercepts but homogeneous slopes can be reformulated as subject to \((N - 1)K\) linear restrictions (non restrictions on \(\alpha_i\)).

\[ H_0^2 : \beta_i = \beta \quad \forall \, i = 1, \ldots, N \]
Definition

Under the assumption $H_0^2$,

$$H_0^2 : \beta_i = \beta \ \forall \ i = 1, .. N$$

the F statistic, denoted $F_2$, and defined by:

$$F_2 = \frac{(RSS_{1,c'} - RSS_1) / [(N - 1) K]}{RSS_1 / [NT - N (K + 1)]}$$

has a Fischer distribution with $(N - 1) K$ et $NT - N (K + 1)$ degrees of freedom under $H_0^2$. $RSS_1$ denotes the residual sum of squares of the model and $RSS_{1,c'}$ the residual sum of squares of the constrained model (model with individual effects):
Specification tests and analysis of covariance

Linear unobserved effects panel data models
Random coefficients models

Specification tests
Application: strikes in OECD

Under $H_0^2$, the residual sum of squares is:

$$RSS_{1,c'} = \sum_{i=1}^{N} S_{yy,i} - \left( \sum_{i=1}^{N} S_{xy,i} \right)' \left( \sum_{i=1}^{N} S_{xx,i} \right)^{-1} \left( \sum_{i=1}^{N} S_{xy,i} \right)$$

$$S_{yy,i} = \sum_{t=1}^{T} (y_{i,t} - \bar{y}_i)^2 \quad \text{with} \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t} \quad \bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}$$

$$S_{xx,i} = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)' (x_{i,t} - \bar{x}_i)$$

$$S_{xy,i} = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)' (y_{i,t} - \bar{y}_i)'$$
Last step

If \( H_0^2 \) is not rejected, one can also apply a conditional test for homogeneous intercepts (\( N - 1 \) linear restrictions).

\[
H_0^3 : \alpha_i = \alpha \ \forall \ i = 1, \ldots, N \ \text{given} \ \beta_i = \beta
\]

- Under the null, the model is homogeneous (pooled) and the restricted residual sum of squares is \( SCR_{1,c} \).
- Under the alternative, the model is \( y_{i,t} = \alpha_i + \beta' x_{i,t} + \epsilon_{i,t} \), and there is \( NT \) – \( N \) degrees of freedom.
Definition

Under the assumption $H_0^3$,

$$H_0^3 : \alpha_i = \alpha \ \forall \ i = 1, .., N \ \text{given} \ \beta_i = \beta$$

the F statistic, denoted $F_3$, and defined by:

$$F_3 = \frac{(RSS_{1,c} - RSS_{1,c'}) / (N - 1)}{RSS_{1,c'} / [N (T - 1) - K]}$$

has a Fischer distribution with $N - 1$ and $N (T - 1) - K$ degrees of freedom under $H_0^2$. $RSS_{1,c'}$ denotes the residual sum of squares of the model with individual effects and $SCR_{1,c}$ the residual sum of squares of the pooled model previously defined.
Specification tests and analysis of covariance
Linear unobserved effects panel data models
Random coefficients models

Application: Strikes in OECD
Example

Let us consider a simple panel regression model for the total number of strikes days in OECD countries. We have a balanced panel data set for 17 countries ($N = 17$) and annual data form 1951 to 1985 ($T = 35$). General idea: evaluate the link between strikes and some macroeconomic factors (inflation, unemployment etc.)

$$s_{i,t} = \alpha_i + \beta_i u_{i,t} + \gamma_i p_{i,t} + \epsilon_{i,t} \quad \forall i = 1, .., 17$$

- $s_{i,t}$ the number of strike days for 1000 workers for the country $i$ at time $t$.
- $u_{i,t}$ the unemployment rate
- $p_{i,t}$, the inflation rate
Figure: Spécification tests with TSP 4.3A

Panel Data Estimation

Balanced data: NI = 17, T = 35, NOB = 595

Total (plain OLS) Estimates:

Dependent variable: SRT

Mean of dependent variable = 305.076
Std. error of regression = 557.258
Std. dev. of dependent var. = 571.637
R-squared = 0.052874
Sum of squared residuals = 0.183838E+09
Adjusted R-squared = 0.049674
Variance of residuals = 310536.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>27.4379</td>
<td>7.53997</td>
<td>3.63899</td>
</tr>
<tr>
<td>P</td>
<td>18.6136</td>
<td>4.98653</td>
<td>3.73277</td>
</tr>
<tr>
<td>C</td>
<td>95.0787</td>
<td>43.1414</td>
<td>2.20389</td>
</tr>
</tbody>
</table>

F test of A,B=Ai,Bi: F(48,544) = 3.8320, P-value = [.0000]
Critical F value for diffuse prior (Leamer, p.114) = 7.6418

Between (OLS on means) Estimates:
Dependent variable: SRT

Mean of dependent variable = 305.076
Std. dev. of dependent var. = 278.196
Sum of squared residuals = 503607.
Variance of residuals = 35972.0

Std. error of regression = 109.60
R-squared = .5933
Adjusted R-squared = .5352

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>80.9542</td>
<td>23.1650</td>
<td>3.49467</td>
</tr>
<tr>
<td>P</td>
<td>59.3882</td>
<td>32.6068</td>
<td>1.82134</td>
</tr>
<tr>
<td>C</td>
<td>-341.546</td>
<td>197.643</td>
<td>-1.72610</td>
</tr>
</tbody>
</table>

WITHIN (fixed effects) Estimates:

Dependent variable: SRT

Sum of squared residuals = .146958E+09
Variance of residuals = 255136.
Std. error of regression = 505.110

R-squared = .2428
Adjusted R-squared = .2192

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>-21.5968</td>
<td>9.19158</td>
<td>-2.34953</td>
</tr>
<tr>
<td>P</td>
<td>16.2729</td>
<td>4.75658</td>
<td>3.42113</td>
</tr>
</tbody>
</table>

F test of Ai,B-Ai,Bi: F(32,544) = 1.1845, P-value = [.2366]
Critical F value for diffuse prior (Leamer, p.114) = 6.9699

F test of A,B-Ai,B: F(16,576) = 9.0342, P-value = [.0000]
Critical F value for diffuse prior (Leamer, p.114) = 6.7476
Variance Components (random effects) Estimates:

\[ \text{VWH} (\text{variance of } U_{it}) = 1.25514 \times 10^6 \]
\[ \text{VBET} (\text{variance of } A_i) = 5.5401 \]

(Computed from small sample formula)
\[ \text{THETA} (0=\text{WITHIN}, 1=\text{TOTAL}) = 0.11628 \]

Dependent variable: \text{SRT}

Sum of squared residuals = 152.560 \times 10^9
R-squared = 0.2140

Variance of residuals = 264862.
Adjusted R-squared = 0.1894

Std. error of regression = 514.647

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>-12.0814</td>
<td>8.78623</td>
<td>-1.37504</td>
</tr>
<tr>
<td>F</td>
<td>16.4247</td>
<td>4.72670</td>
<td>3.47489</td>
</tr>
<tr>
<td>C</td>
<td>240.622</td>
<td>71.9106</td>
<td>3.45737</td>
</tr>
</tbody>
</table>

Hausman test of H0: RE vs. FE: CHISQ(2) = 13.924, P-value = [0.0009]
Conclusion of section 1.

Fact

*It is possible to test the heterogeneity / homogeneity of the parameters under some specific assumptions* (normality of residuals). More generally, the assumption of heterogeneity / homogeneity of the parameters (slope coefficients and constants) has to be evaluated with an economic interpretation.

Example

Example: It is reasonable to assume that the slope parameters of the production function are the same across countries? What does it imply? Should I impose a common mean for the TFP for France and Germany?
Section 2.
Linear unobserved/individual effects panel data models
2.1. Definitions
We now consider a model with individual effects and common slope parameters.

**Definition**

A linear unobserved / individual effects panel data model is defined as follows:

\[ y_{it} = \alpha_i + \beta' x_{it} + \epsilon_{it} \]

where \( \forall i = 1, \ldots, N, \forall t = 1, \ldots, T \)

\( \alpha_i \) and \( \beta = (\beta_1, \beta_2, \ldots, \beta_K) \) are \( 1 \times 1 \) and \( 1 \times K \) vectors of parameters,

\( x_{it} = (x_{1it}, \ldots, x_{Kit}) \) is a \( 1 \times K \) vector of exogenous variables,

\( \epsilon_{it} \) is the error term, assumed to be i.i.d., with \( \forall i = 1, \ldots, N, \forall t = 1, \ldots, T \)

\[ E(\epsilon_{it}) = 0 \quad E(\epsilon_{it}^2) = \sigma^2 \]
Linear unobserved effects panel data models

**Definition**
There are many names for $\alpha_i$: (1) unobserved effects, (2) individual effects, (3) unobserved component, (4) latent variable (for random effects models), (5) individual heterogeneity.

**Definition**
The $\varepsilon_{it}$ are called the **idiosyncratic errors** or **idiosyncratic disturbances** because these change across $t$ as well as across $i$. 
Linear unobserved effects panel data models

Writing in vector form. Let us denote

\[
y_i(\mathbf{T},1) = \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{pmatrix} \quad \quad X_i(\mathbf{T},K) = \begin{pmatrix} x_{1,i,1} & x_{2,i,1} & \cdots & x_{K,i,1} \\ x_{1,i,2} & x_{2,i,2} & \cdots & x_{K,i,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,i,T} & x_{2,i,T} & \cdots & x_{K,i,T} \end{pmatrix}
\]

Let us denote \( e \) a unit vector and \( \varepsilon_i \) the vector of errors:

\[
e(\mathbf{T},1) = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \quad \varepsilon_i(\mathbf{T},1) = \begin{pmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,T} \end{pmatrix}
\]
Linear unobserved effects panel data models

**Definition**

For a given individual $i$, the linear unobserved / individual effects panel data model is defined as follows:

$$y_i = e\alpha_i + X_i\beta + \varepsilon_i \quad \forall i = 1, \ldots, N$$

$$E(\varepsilon_i) = 0$$

$$E(\varepsilon_i\varepsilon_i') = \sigma^2_\varepsilon I_T$$

$$E(\varepsilon_i\varepsilon_j') = 0 \quad (T, T) \text{ if } i \neq j$$
Linear unobserved effects panel data models

Example

Let us consider the case of a Cobb Douglas production function in log, as defined previously, for the case $T = 3$. We have:

$$y_{i,t} = \alpha_i + \beta_k k_{i,t} + \beta_n n_{i,t} + \varepsilon_{i,t} \quad \forall i, \forall t \in [1, 3]$$

or in a vectorial form for a country $i$:

$$\begin{pmatrix} y_{i,1} \\ y_{i,2} \\ y_{i,3} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \alpha_i + \begin{pmatrix} k_{i,1} & n_{i,1} \\ k_{i,2} & n_{i,2} \\ k_{i,3} & n_{i,3} \end{pmatrix} \begin{pmatrix} \beta_k \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \varepsilon_{i,3} \end{pmatrix}$$
Linear unobserved effects panel data models

It is also possible to stackle all this vectors/matrix as follows

\[ Y = \tilde{e}\tilde{\alpha} + X\beta + \varepsilon \]

\[
Y_{(TN,1)} = \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_N
\end{pmatrix}, \\
X_{(TN,K)} = \begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_N
\end{pmatrix}, \\
\varepsilon_{(TN,1)} = \begin{pmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\vdots \\
\varepsilon_N
\end{pmatrix}
\]

where \(0_T\) is the null vector \((T, 1)\).

\[
\tilde{e}_{(TN,N)} = \begin{pmatrix}
e & 0_T & \ldots & 0_T \\
0_T & e & \ldots & 0_T \\
\vdots & \vdots & \ddots & \vdots \\
0_T & 0_T & \ldots & e
\end{pmatrix}, \\
\tilde{\alpha}_{(N,1)} = \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_N
\end{pmatrix}
\]

C. Hurlin

Panel Data Econometrics
Linear unobserved effects panel data models

Example

Consider the case of the production function with $T = 3$ and $N = 2$

$$\begin{pmatrix} y_{1,1} \\ y_{1,2} \\ y_{1,3} \\ y_{2,1} \\ y_{2,2} \\ y_{2,3} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} k_{11} & n_{11} \\ k_{12} & n_{12} \\ k_{13} & n_{13} \\ k_{21} & n_{21} \\ k_{22} & n_{22} \\ k_{3} & n_{23} \end{pmatrix} \begin{pmatrix} \beta_k \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{1} \\ \epsilon_{2,1} \\ \epsilon_{2,2} \\ \epsilon_{2} \end{pmatrix}$$

$$\iff Y = \tilde{e}\bar{\alpha} + X\beta + \epsilon$$
2.2. Fixed and random effects
Especially in methodological papers, but also in applications, one often sees a discussion about whether $\alpha_i$ will be treated as a random effect or a fixed effect.

**Definition**

In the traditional approach to panel data models, $\alpha_i$ is called a “random effect” when it is treated as a random variable and a “fixed effect” when it is treated as a parameter to be estimated for each cross section observation $i$. 
Fact

For Wooldridge (2003), these discussions about whether the $\alpha_i$ should be treated as random variables or as parameters to be estimated are wrongheaded for microeconometric panel data applications. With a large number of random draws from the cross section, it almost always makes sense to treat the unobserved effects, $\alpha_i$, as random draws from the population, along with $y_{it}$ and $x_{it}$. This approach is certainly appropriate from an omitted variables or neglected heterogeneity perspective.
Fact

As suggested by Mundlak (1978), the key issue involving $\alpha_i$ is whether or not it is uncorrelated with the observed explanatory variables $x_{it}$, for $t = 1, ..., T$.

In modern econometric parlance, “random effect” is synonymous with zero correlation between the observed explanatory variables and the unobserved effect:

\[
cov(x_{it}, \alpha_i) = 0, \quad t = 1, \ldots, T
\]
Actually, a stronger conditional mean independence assumption,

$$E (\alpha_i \mid x_{i1}, \ldots, x_{iT}) = 0$$

will be needed to fully justify statistical inference. In applied papers, when $\alpha_i$ is referred to as, say, an “individual random effect,” then $\alpha_i$ is probably being assumed to be uncorrelated with the $x_{it}$.
In microeconometric applications, the term “fixed effect” does not usually mean that $\alpha_i$ is being treated as nonrandom; rather, it means that one is allowing for arbitrary correlation between the unobserved effect $\alpha_i$ and the observed explanatory variables $x_{it}$. 
Wooldridge (2003) avoids referring to $\alpha_i$ as a random effect or a fixed effect. Instead, we will refer to $\alpha_i$ as unobserved effect, unobserved heterogeneity, and so on. Nevertheless, later we will label two different estimation methods random effects estimation and fixed effects estimation. This terminology is so ingrained that it is pointless to try to change it now.
Linear unobserved effects panel data models

Example (Production function)

Let us consider the simple example of the Cobb Douglass production function.

\[ y_{i,t} = \beta_i k_{i,t} + \gamma_i n_{i,t} + \alpha_i + v_{i,t} \]

In this case, \( \alpha_i \) corresponds to the unobserved effect on TFP due to country specific omitted factor (climate, institutions, organization etc..). In this case, we might expect that the the level of factors are positively correlated with this component of TFP: the more a country is productive, the more it invests in capital for instance.

\[ \text{cov} (\alpha_i, k_{i,t}) > 0 \quad \text{cov} (\alpha_i, n_{i,t}) > 0 \]
Example (Patents and R&D)

Hausman, Hall, and Griliches (1984) estimate (nonlinear) distributed lag models to study the relationship between patents awarded to a firm and current and past levels of R&D spending. A linear version of their model is:

\[ \text{patents}_{it} = \theta_t + z_{it} \gamma + \delta_0 RD_{it} + \delta_1 RD_{i,t-1} + \ldots + \delta_5 RD_{i,t-5} + \alpha_i + v_{i,t} \]

where \( RD_{it} \) is spending on R&D for firm \( i \) at time \( t \) and \( z_{it} \) contains other explanatory variables. \( \alpha_i \) is a firm heterogeneity term that may influence \( \text{patents}_{it} \) and that may be correlated with current, past, and future R&D expenditures.

\[ \text{cov} (\alpha_i, RD_{i,t-k}) \neq 0 \quad \forall k \]
The Hausman’s test (1978), is a test of the null hypothesis

\[ \text{cov}(x_{it}, \alpha_i) = 0, \quad t = 1, \ldots, T \]

and is generally presented as test of specification (fixed or random) of the unobserved effects (see later, section 5.).
2.3. Fixed effects methods
Let us consider the following model:

\[ y_i = e\alpha_i + X_i\beta + \varepsilon_i \quad \forall i = 1, \ldots, N \]

where \( \alpha_i \) is a constant term, \( \beta' = (\beta_1, \beta_2, \ldots, \beta_K) \in \mathbb{R}^K \) and:

\[
\begin{align*}
    y_i & = \begin{pmatrix} \varepsilon_{i,1} & \varepsilon_{i,2} & \ldots & \varepsilon_{i,T} \end{pmatrix}' \\
    \varepsilon_i & = \begin{pmatrix} \varepsilon_{i,1} & \varepsilon_{i,2} & \ldots & \varepsilon_{i,T} \end{pmatrix}' \\
    e & = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}' \\
    X_i & = \begin{pmatrix} x_{1,1,i} & x_{2,1,i} & \ldots & x_{K,1,i} \\
                   x_{1,2,i} & x_{2,2,i} & \ldots & x_{K,2,i} \\
                   \vdots & \vdots & \ddots & \vdots \\
                   x_{1,T,i} & x_{2,T,i} & \ldots & x_{K,T,i} \end{pmatrix}
\end{align*}
\]
Linear unobserved effects panel data models

Assumptions (H1) Let us assume that errors terms $\varepsilon_{i,t}$ are i.i.d. $\forall i \in [1, N], \forall t \in [1, T]$ with:

- $E(\varepsilon_{i,t}) = 0$
- $E(\varepsilon_{i,t}\varepsilon_{i,s}) = \begin{cases} \sigma_{\varepsilon}^2 & t = s \\ 0 & \forall t \neq s \end{cases}$, or $E(\varepsilon_i\varepsilon'_i) = \sigma_{\varepsilon}^2 I_T$

where $I_T$ denotes the identity matrix $(T, T)$.

- $E(\varepsilon_{i,t}\varepsilon_{j,s}) = 0, \forall j \neq i, \forall (t, s)$, or $E(\varepsilon_i\varepsilon'_j) = 0$

where $0$ denotes the null matrix $(T, T)$. 
Linear unobserved effects panel data models

**Theorem**

Under assumption $H_1$, the ordinary-least-squares (OLS) estimator of $\beta$ is the best linear unbiased estimator (BLUE).

**Definition**

In this context, the OLS estimator $\hat{\beta}$ is called the least-squares dummy-variable (LSDV) of Fixed Effect (FE) estimator, because the observed values of the variable for the coefficient $\alpha_i$ takes the form of dummy variables.
OLS estimators of $\alpha_i$ and $\beta$ and are obtained by minimizing

\[
\left\{ \widehat{\alpha}_i, \widehat{\beta}_{LSDV} \right\} \arg \min_{\{\alpha_i, \beta\}_{i=1}^N} S = \sum_{i=1}^N \varepsilon_i' \varepsilon_i \\
= \sum_{i=1}^N (y_i - e\alpha_i - X_i\beta)' (y_i - e\alpha_i - X_i\beta)
\]
Linear unobserved effects panel data models

FOC1 (with respect to $\alpha_i$):

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_{LSDV}'\bar{x}_i$$

with

$$\bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t} \quad \bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}$$
Given the second FOC (with respect to $\beta$) and the previous result, we have:

**Definition**

Under assumption $H_1$, the fixed effect estimator or LSDV estimator of parameter $\beta$ is defined by:

$$
\hat{\beta}_{LSDV} = \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i) (x_{i,t} - \bar{x}_i)' \right]^{-1} \\
\left[ \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i) (y_{i,t} - \bar{y}_i) \right]
$$
1. The computational procedure for estimating the slope parameters in this model does not require that the dummy variables for the individual (and/or time) effects actually be included in the matrix of explanatory variables.

2. We need only find the means of time-series observations separately for each cross-sectional unit, transform the observed variables by subtracting out the appropriate time-series means, and then apply the least squares method to the transformed data.
The foregoing procedure is equivalent to premultiplying the $i^{th}$ equation

$$y_i = e\alpha_i + X_i \beta + \varepsilon_i$$

by a $T \times T$ idempotent (covariance) transformation matrix (within operator)

$$Q = I_T - \frac{1}{T} ee'$$

to “sweep out” the individual effect $\alpha_i$ so that individual observations are measured as deviations from individual means (over time).
Linear unobserved effects panel data models

$Qy_i$ and $QX_i$ correspond to the observations are measured as deviations from individual means:

$$Qy_i = \left( I_T - \frac{1}{T} ee' \right) y_i$$

$$= y_i - e \left( \frac{1}{T} e' y_i \right)$$

$$= \begin{pmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,T} \end{pmatrix} - \begin{pmatrix} \frac{1}{T} \sum_{t=1}^{T} y_{i,t} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$
Linear unobserved effects panel data models

\[ QX_i = X_i - \frac{1}{T} ee' X_i \]

\[ = \begin{pmatrix} x_{1,i,1} & x_{2,i,1} & \ldots & x_{K,i,1} \\ x_{1,i,2} & x_{2,i,2} & \ldots & x_{K,i,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,i,T} & x_{2,i,T} & \ldots & x_{K,i,T} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} - \frac{1}{T} \begin{pmatrix} \sum_{t=1}^T x_{1,i,t} \\ \sum_{t=1}^T x_{2,i,t} \\ \vdots \\ \sum_{t=1}^T x_{K,i,t} \end{pmatrix} \]
Linear unobserved effects panel data models

Finally, when the transformation $Q$ is applied to a vector of constant (or a time invariant variable), it leads to a null vector.

$$Qe = \left( I_T - \frac{1}{T} ee' \right) e$$

$$= e - \frac{1}{T} ee'e$$

$$= e - e = 0$$

since

$$e'e = \begin{pmatrix} 1 & \ldots & 1 \\ 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = T$$
So, we have:

\[ y_i = e\alpha_i + X_i\beta + \varepsilon_i \]

\[ \iff Qy_i = Qe\alpha_i + QX_i\beta + Q\varepsilon_i \]

\[ \iff Qy_i = QX_i\beta + Q\varepsilon_i \]
Linear unobserved effects panel data models

Definition

Under assumption $H_1$, the fixed effect estimator or LSDV estimator of parameter $\beta$ is defined by:

$$\hat{\beta}_{LSDV} = \left[ \sum_{i=1}^{N} X'_i Q X_i \right]^{-1} \left[ \sum_{i=1}^{N} X'_i Q y_i \right]$$

where

$$Q = I_T - \frac{1}{T} ee'$$
Linear unobserved effects panel data models

Example

Let us consider a simple panel regression model for the total number of strikes days in OECD countries. We have a balanced panel data set for 17 countries ($N = 17$) and annual data from 1951 to 1985 ($T = 35$). General idea: evaluate the link between strikes and some macroeconomic factors (inflation, unemployment etc..)

$$s_{i,t} = \alpha_i + \beta_i u_{i,t} + \gamma_i p_{i,t} + \epsilon_{i,t} \quad \forall i = 1, \ldots, 17$$

- $s_{i,t}$ the number of strike days for 1000 workers for the country $i$ at time $t$.
- $u_{i,t}$ the unemployment rate
- $p_{i,t}$, the inflation rate
**PANEL DATA ESTIMATION**

**Balanced data:** NI= 17, T= 35, NOB= 595

**WITHIN (fixed effects) Estimates:**

**Dependent variable:** SRT

\[
\text{Sum of squared residuals} = .146958E+09 \\
\text{Variance of residuals} = 255136. \\
\text{Std. error of regression} = 505.110
\]

<table>
<thead>
<tr>
<th>Estimated Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>-21.5968</td>
<td>9.19158</td>
<td>-2.34963</td>
</tr>
<tr>
<td>P</td>
<td>16.2729</td>
<td>4.75658</td>
<td>3.42113</td>
</tr>
</tbody>
</table>

R\-squared = .242875

Adjusted R\-squared = .219215
Method of estimation = Ordinary Least Squares

Dependent variable: SRTC  
Number of observations: 595

Mean of dependent variable = -.159906E-07  
Std. dev. of dependent var. = 503.791  
Sum of squared residuals = .146958E+09  
Variance of residuals = 247822.  
Std. error of regression = 497.817  
R-squared = .023220  
Adjusted R-squared = .023576  
Durbin-Watson statistic = 1.98516  
F-statistic (zero slopes) = 15.3421  
Schwarz Bayes. Info. Crit. = 12.4386  
Log of likelihood function = -4530.36

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>UC</td>
<td>-21.5968</td>
<td>9.05887</td>
<td>-2.38405</td>
</tr>
<tr>
<td>PC</td>
<td>16.2729</td>
<td>4.68791</td>
<td>3.47125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>355.49000</td>
<td>-2.58225</td>
<td>801.98611</td>
<td>211.05529</td>
</tr>
<tr>
<td>387.83512</td>
<td>339.06634</td>
<td>57.54147</td>
<td>582.61959</td>
</tr>
<tr>
<td>1012.66300</td>
<td>114.22273</td>
<td>27.49515</td>
<td>164.23965</td>
</tr>
<tr>
<td>14.43655</td>
<td>14.93684</td>
<td>-44.85353</td>
<td>284.45232</td>
</tr>
</tbody>
</table>

C. Hurlin  
Panel Data Econometrics
Theorem

The LSDV estimator $\hat{\beta}$ is unbiased and consistent when either $N$ or $T$ or both tend to infinity. Its variance–covariance matrix is:

$$V \left( \hat{\beta}_{LSDV} \right) = \sigma^2 \left[ \sum_{i=1}^{N} X_i' Q X_i \right]^{-1}$$

$$\hat{\beta}_{LSDV} \overset{p}{\rightarrow} \beta \quad \text{as} \quad NT \rightarrow \infty$$
However, the estimator for the intercept, $\hat{\alpha}_i$, although unbiased, is consistent only when $T \to \infty$.

$$\hat{\alpha}_i \xrightarrow{p} \alpha_i$$
2.4. Random effects methods
The random specification of unobserved effects corresponds to a particular case of variance-component or error-component model, in which the residual is assumed to consist of three components

\[ y_{i,t} = \beta' x_{i,t} + \varepsilon_{i,t} \quad \forall \ i \forall \ t \]

\[ \varepsilon_{i,t} = \alpha_i + \lambda_t + \upsilon_{i,t} \]
Assumptions (H2) Let us assume that error terms \( \varepsilon_{i,t} = \alpha_i + \lambda_t + \nu_{i,t} \) are i.i.d. with
\( \forall i = 1, \ldots, N, \forall t = 1, \ldots, T \)
- \( E(\alpha_i) = E(\lambda_t) = E(\nu_{i,t}) = 0 \)
- \( E(\alpha_i\lambda_t) = E(\lambda_t\nu_{i,t}) = E(\alpha_i\nu_{i,t}) = 0 \)
- \( E(\alpha_i\alpha_j) = \begin{cases} \sigma_{\alpha}^2 & i = j \\ 0 & \forall i \neq j \end{cases} \)
- \( E(\lambda_t\lambda_s) = \begin{cases} \sigma_{\lambda}^2 & t = s \\ 0 & \forall t \neq s \end{cases} \)
- \( E(\nu_{i,t}\nu_{j,s}) = \begin{cases} \sigma_{\nu}^2 & t = s, i = j \\ 0 & \forall t \neq s, \forall i \neq j \end{cases} \)
- \( E(\alpha_i\lambda_{i,t}) = E(\lambda_{t}\lambda_{i,t}) = E(\nu_{i,t}\lambda_{i,t}) = 0 \)
As suggested by Wooldridge (2001), the "fixed effect" specification can be viewed as a case in which $\alpha_i$ is a random parameter with $\text{cov}(\alpha_i, x_{i,t}') \neq 0$, whereas the "random effect model" corresponds to the situation in which $\text{cov}(\alpha_i, x_{i,t}') = 0$.

The variance of $y_{it}$ conditional on $x_{it}$ is the sum of three components:

$$\sigma_y^2 = \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_v^2$$
Definition

Sometimes, the individual effects $\alpha_i^*$ are supposed to have a non-zero mean, with $E(\alpha_i) = \mu$, then we can defined a individual effects $\alpha_i$ with zero mean. A linear unobserved effects panel data models is then defined as follows:

$$y_i = e\mu + X_i\beta + \epsilon_i$$

$$\epsilon_i \sim e^{\alpha_i} + \nu_i$$

$(T,1)$ $(T,1)(T,1)$ $(T,1)$
The vectorial expression of the individual effects model is then:

\[
y_i = \tilde{X}_i \gamma + \varepsilon_i
\]

\[
\varepsilon_i = e \alpha_i + \nu_i
\]

with

\[
\tilde{X}_i = (e : X_i) \quad \text{and} \quad \gamma' = (\mu : \beta')
\]
Definition

Under assumptions $H_2$, the variance-covariance matrix of $\varepsilon_i$ is equal to:

$$V = E(\varepsilon_i\varepsilon'_i) = E[(\alpha_i e + v_i)(\alpha_i e + v_i)'] = \sigma^2_\alpha ee' + \sigma^2_v I_T$$

Its inverse is:

$$V^{-1} = \frac{1}{\sigma^2_v} \left[ I_T - \left( \frac{\sigma^2_\alpha}{\sigma^2_v + T\sigma^2_\alpha} \right) ee' \right]$$
The presence of $\alpha_i$ produces a correlation among residuals of the same crosssectional unit, though the residuals from different cross-sectional units are independent.

$$V = E (\varepsilon_i \varepsilon_i') = \sigma_\alpha^2 \varepsilon \varepsilon' + \sigma_v^2 I_T$$

$$V = \begin{pmatrix}
\sigma_\alpha^2 + \sigma_v^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 \\
\sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_v^2 & \ldots & \sigma_\alpha^2 \\
\ldots & \ldots & \sigma_\alpha^2 & \sigma_v^2 \\
\sigma_\alpha^2 & \sigma_\alpha^2 & \ldots & \sigma_\alpha^2 + \sigma_v^2
\end{pmatrix}$$
However, regardless of whether the $\alpha_i$ are treated as fixed or as random, the individual-specific effects for a given sample can be swept out by the idempotent (covariance) transformation matrix $Q$

$$Qy_i = Qe\mu + QX_i\beta + Qe\alpha_i + Qv_i$$

Since $Qe = (I_T - T^{-1}ee')e = 0$, we have

$$Qy_i = Qe\mu + QX_i\beta + Qv_i$$
Under assumptions $H_2$, when $\alpha_i$ are treated as random, the LSDV estimator is unbiased and consistent either $N$ or $T$ or both tend to infinity. However, whereas the LSDV is the BLUE under the assumption that $\alpha_i$ are fixed constants, it is not the BLUE in finite samples when $\alpha_i$ are assumed random. The BLUE in the latter case is the generalized-least-squares (GLS) estimator.
Moreover, if the explanatory variables contain some time-invariant variables $z_i$, their coefficients cannot be estimated by LSDV, because the covariance transformation eliminates $z_i$. 
Let us consider the model

\[ y_i = \tilde{X}_i \gamma + \varepsilon_i \quad \forall i = 1, \ldots, N \]

where \( \varepsilon_i = \alpha_i e + \nu_i \), \( \tilde{X}_i = (e, X_i) \) and \( \gamma' = (\mu, \beta') \). The variance covariance matrix \( V = E(\varepsilon_i \varepsilon_i') \) is known.

**Definition**

*If the variance covariance matrix \( V \) is known, the GLS estimator of the \( \gamma \) vector, denoted \( \hat{\gamma}_{GLS} \), is defined by:

\[
\hat{\gamma}_{GLS} = \left[ \sum_{i=1}^{N} \tilde{X}_i' V^{-1} \tilde{X}_i \right]^{-1} \left[ \sum_{i=1}^{N} \tilde{X}_i' V^{-1} y_i \right]
\]

Under assumptions \( H_2 \), this estimator is BLUE.*
Definition

Following Maddala (1971), we write $V$ as:

$$V^{-1} = \frac{1}{\sigma_v^2} \left[ Q + \psi \frac{1}{T} ee' \right]$$

where $Q = (I_T - ee' / T)$ and where parameter $\psi$ is defined by:

$$\psi = \left( \frac{\sigma_v^2}{\sigma_v^2 + T \sigma_\alpha^2} \right)$$
Given this definition of $V^{-1}$, we have:

$$
\hat{\gamma}_{GLS} = \left[ \sum_{i=1}^{N} \tilde{X}_i' \left( Q + \psi \frac{1}{T} ee' \right) \tilde{X}_i \right]^{-1} \left[ \sum_{i=1}^{N} \tilde{X}_i' \left( Q + \psi \frac{1}{T} ee' \right) y_i \right]
$$

$$
\hat{\gamma}_{GLS} = \left[ \sum_{i=1}^{N} \tilde{X}_i' Q \tilde{X}_i + \psi \frac{1}{T} \sum_{i=1}^{N} \tilde{X}_i' ee' \tilde{X}_i \right]^{-1} \left[ \sum_{i=1}^{N} \tilde{X}_i' Q y_i + \psi \frac{1}{T} \sum_{i=1}^{N} \tilde{X}_i' ee' y_i \right]
$$

with $\tilde{X}_i = (e \ X_i)$ and $\gamma' = (\mu \ \beta')$
Linear unobserved effects panel data models

\[
\begin{bmatrix}
\hat{\mu}_{GLS} \\
\hat{\beta}_{GLS}
\end{bmatrix}
= \begin{bmatrix}
\psi NT & \psi T \sum_{i=1}^{N} \bar{x}_i' \\
\psi T \sum_{i=1}^{N} \bar{x}_i & \sum_{i=1}^{N} X_i' QX_i + \psi T \sum_{i=1}^{N} \bar{x}_i \bar{x}_i' \\
\psi NT \bar{y} & \\
\sum_{i=1}^{N} X_i' Qy_i + \psi T \sum_{i=1}^{N} \bar{x}_i \bar{y}_i
\end{bmatrix}^{-1}
\]

Using the formula of the partitioned inverse, we can derive \( \hat{\beta}_{GLS} \).
Definition

If the variance covariance matrix $V$ is known, the GLS estimator of vector $\beta$ is defined by:

$$
\hat{\beta}_{GLS} = \left[ \frac{1}{T} \sum_{i=1}^{N} X_i' Q X_i + \psi \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \right]^{-1} 
\left[ \frac{1}{T} \sum_{i=1}^{N} X_i' Q y_i + \psi \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{y}_i - \bar{y}) \right]
$$

where $\psi$ is a constant defined by $\psi = \sigma_v^2 \left( \sigma_v^2 + T \sigma_\alpha^2 \right)^{-1}$
This estimator can be expressed as a weighted average of the LSDV (OLS) estimator and the between estimator.

**Definition**

The between-group estimator or between estimator, denoted $\hat{\beta}_{BE}$, corresponds to the OLS estimator obtained in the model:

$$\bar{y}_i = c + \beta' \bar{x}_i + \varepsilon_i \quad \forall \ i = 1, .., N$$

$$\hat{\beta}_{BE} = \left[ \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \right]^{-1} \left[ \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{y}_i - \bar{y}) \right]$$

The estimator $\hat{\beta}_{BE}$ is called the between-group estimator because it ignores variation within the group.
Definition

The pooled estimator, denoted \( \hat{\beta}_{pooled} \), corresponds to the OLS estimator obtained in the model:

\[
y_{it} = \alpha + \beta' x_{it} + \varepsilon_{it} \quad \forall \ i = 1, \ldots, N \quad \forall \ t = 1, \ldots, T
\]

\[
\hat{\beta}_{pooled} = \left[ \sum_{t=1}^{T} \sum_{i=1}^{N} (x_{i,t} - \bar{x})(\bar{x}_i - \bar{x})' \right]^{-1} \left[ \sum_{t=1}^{T} \sum_{i=1}^{N} (x_{it} - \bar{x})(y_{it} - \bar{y}) \right]
\]
Theorem

Under assumptions $H_2$, the GLS estimator $\hat{\beta}_{GLS}$ is a weighted average of the between-group $\hat{\beta}_{BE}$ and the within-group (LSDV) estimators $\hat{\beta}_{LSDV}$.

$$\hat{\beta}_{GLS} = \Delta \hat{\beta}_{BE} + (I_K - \Delta) \hat{\beta}_{LSDV}$$

where $\Delta$ denotes a weighted matrix defined by:

$$\Delta = \psi^T \left[ \sum_{i=1}^{N} X'_i Q X_i + \psi^T \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \right]^{-1} \left[ \sum_{i=1}^{N} (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \right]$$
If $\psi \to 1$, then GLS converges to the OLS pooled estimator.

$$\hat{\beta}_{GLS} \xrightarrow[p]{\psi \to 1} \hat{\beta}_{pooled}$$

If $\psi \to 0$, the GLS estimator converges to LSDV estimator.

$$\hat{\beta}_{GLS} \xrightarrow[p]{\psi \to 0} \hat{\beta}_{LSDV}$$
Fact

The constant $\psi$ measures the weight given to the between-group variation.

In the LSDV (or fixed-effects model) procedure, this source of variation is completely ignored.

The OLS procedure corresponds to $\psi = 1$. The between-group and within-group variations are just added up.
Fact

The procedure of treating $\alpha_i$ as random provides a solution intermediate between treating them all as different and treating them all as equal.
Linear unobserved effects panel data models

Given the definition of $\psi$, we have:

$$
\psi = \left( \frac{\sigma^2_v}{\sigma^2_v + T\sigma^2_{\alpha}} \right) \quad \lim_{T \to \infty} \psi = 0
$$

Fact

When $T$ tends to infinity, the GLS estimator converges to the LSDV estimator:

$$
\hat{\beta}_{GLS} \xrightarrow{T \to \infty} \hat{\beta}_{LSDV}
$$

This is because when $T \to \infty$, we have an infinite number of observations for each $i$. Therefore, we can consider each $\alpha_i$ as a random variable which has been drawn once and forever, so that for each $i$ we can pretend that they are just like fixed parameters.
Fact

Computation of the GLS estimator can be simplified by introducing a transformation matrix $P$ such that

$$P = \left[ I_T - \left( 1 - \psi^{1/2} \right) \left( 1 / T \right) ee' \right]$$

We have

$$V^{-1} = \frac{1}{\sigma_v^2} P' P$$

Premultiplying the model by the transformation matrix $P$, we obtain the GLS estimator by applying the least-squares method to the transformed model (Theil (1971, Chapter 6)).
This is equivalent to first transforming the data by subtracting a fraction \((1 - \psi^{1/2})\) of individual means \(\bar{y}_i\) and \(\bar{x}_i\) from their corresponding \(y_{it}\) and \(x_{it}\), then regressing \(y_{it} - (1 - \psi^{1/2}) \bar{y}_i\) on a constant and \(x_{it} - (1 - \psi^{1/2}) \bar{x}_i\).
We can show that:

\[
\text{var} \left( \hat{\beta}_{GLS} \right) = \sigma_v^2 \left[ \sum_{i=1}^{N} X_i' Q X_i + \psi T \sum_{i=1}^{N} (\overline{x}_i - \overline{x}) (\overline{x}_i - \overline{x})' \right]^{-1}
\]

Because \( \psi > 0 \), we see immediately that the difference between the covariance matrices of \( \hat{\beta}_{LSDV} \) and \( \hat{\beta}_{GLS} \) is a positive semidefinite matrix. For \( K = 1 \), we have:

\[
\text{var} \left( \hat{\beta}_{GLS} \right) \leq \text{var} \left( \hat{\beta}_{LSDV} \right)
\]
Linear unobserved effects panel data models

Definition

If the variance components $\sigma^2_\varepsilon$ and $\sigma^2_\alpha$ are unknown, we can use two-step GLS estimation.

1. In the first step, we estimate the variance components using some consistent estimators.
2. In the second step, we substitute their estimated values into

$$\hat{\gamma}_{GLS} = \left[ \sum_{i=1}^{N} \tilde{X}_i \tilde{V}^{-1} \tilde{X}_i \right]^{-1} \left[ \sum_{i=1}^{N} \tilde{X}_i \tilde{V}^{-1} y_i \right]$$

or its equivalent form.
Linear unobserved effects panel data models

Lemma

When the sample size is large (in the sense of either $N \to \infty$, or $T \to \infty$), the two-step GLS estimator will have the same asymptotic efficiency as the GLS procedure with known variance components.

Lemma

Even for moderate sample size [for $T \geq 3$, $N - (K + 1) \geq 9$; for $T \geq 2$, $N - (K + 1) \geq 10$], the two-step procedure is still more efficient than the covariance (or within-group) estimator in the sense that the difference between the covariance matrices of the covariance estimator and the two-step estimator is nonnegative definite.
Noting that \( \bar{y}_i = \alpha_i + \beta' \bar{x}_i + \bar{\varepsilon}_i \) and 
\[
( y_{it} - \bar{y}_i ) = ( x_{it} - \bar{x}_i ) + ( v_{it} - \bar{v}_i ),
\]
we can use the within- and between-group residuals to estimate \( \sigma^2_{\varepsilon} \) and \( \sigma^2_{\alpha} \) by

\[
\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left[ ( y_{i,t} - \bar{y}_i ) - \hat{\beta}'_{LSDV} ( x_{i,t} - \bar{x}_i ) \right]^2}{N ( T - 1 ) - K}
\]

\[
\hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^{N} \left( \bar{y}_i - \hat{\beta}'_{LSDV} \bar{x}_i \right)^2}{N - K - 1} - \hat{\sigma}_v^2
\]
Then, we have an estimate of $\psi$ and $V^{-1}$

$$\hat{\psi} = \left( \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + T\hat{\sigma}_\alpha^2} \right)$$

$$\hat{V}^{-1} = \frac{1}{\hat{\sigma}_v^2} \left[ Q + \hat{\psi} \frac{1}{T} ee' \right]$$
Example

Let us consider a simple panel regression model for the total number of strikes days in OECD countries. We have a balanced panel data set for 17 countries ($N = 17$) and annual data form 1951 to 1985 ($T = 35$).

$$s_{i,t} = \alpha_i + \beta u_{i,t} + \gamma p_{i,t} + \epsilon_{i,t} \quad \forall i = 1, \ldots, 17$$

- $s_{i,t}$ the number of strike days for 1000 workers for the country $i$ at time $t$.
- $u_{i,t}$ the unemployment rate
- $p_{i,t}$ the inflation rate
**Figure: Random effects method**

**PANEL DATA ESTIMATION**

Balanced data: NI= 17, T= 35, NOB= 595

Variance Components (random effects) Estimates:

\[ V_{WITH} \text{ (variance of } U_t) = 0.25514E+06 \]
\[ V_{BET} \text{ (variance of } A_i) = 55401. \]
(computed from small sample formula)

THETA (0=WITHIN, 1=TOTAL) = 0.11628

Dependent variable: SRT

Sum of squared residuals = .152560E+09 \hspace{1cm} \text{R-squared} = .214013

Variance of residuals = 264362. \hspace{1cm} \text{Adjusted R-squared} = .189450

Std. error of regression = 514.647

<table>
<thead>
<tr>
<th>Estimated Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>-12.0814</td>
<td>8.78623</td>
<td>-1.37504</td>
</tr>
<tr>
<td>P</td>
<td>16.4247</td>
<td>4.72670</td>
<td>3.47489</td>
</tr>
<tr>
<td>C</td>
<td>248.622</td>
<td>71.9106</td>
<td>3.45737</td>
</tr>
</tbody>
</table>
2.5. Fixed or random methods?
Linear unobserved effects panel data models

Fact

Whether to treat the effects as fixed or random makes no difference when $T$ is large, because both the LSDV estimator and the generalized least-squares estimator become the same estimator:

$$\widehat{\beta}_{GLS} \xrightarrow{T \to \infty} \widehat{\beta}_{LSDV}$$
When $T$ is finite and $N$ is large, whether to treat the effects as fixed or random is not an easy question to answer. It can make a surprising amount of difference in the estimates of the parameters.
Example

For example, Hausman (1978) found that using a fixed-effects specification produced significantly different results from a random-effects specification when estimating a wage equation using a sample of 629 high school graduates followed over six years by the Michigan income dynamics study. The explanatory variables in the Hausman wage equation include a piecewise-linear representation of age, the presence of unemployment or poor health in the previous year, and dummy variables for self-employment, living in the South, or living in a rural area.
### Table 3.3. Wage equations (dependent variable: log wage)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed effects</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age 1 (20–35)</td>
<td>0.0557</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>2. Age 2 (35–45)</td>
<td>0.0351</td>
<td>0.0092</td>
</tr>
<tr>
<td></td>
<td>(0.0051)</td>
<td>(0.0036)</td>
</tr>
<tr>
<td>3. Age 3 (45–55)</td>
<td>0.0209</td>
<td>-0.0007</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0042)</td>
</tr>
<tr>
<td>4. Age 4 (55–65)</td>
<td>0.0209</td>
<td>-0.0097</td>
</tr>
<tr>
<td></td>
<td>(0.0078)</td>
<td>(0.0060)</td>
</tr>
<tr>
<td>5. Age 5 (65–)</td>
<td>-0.0171</td>
<td>-0.0423</td>
</tr>
<tr>
<td></td>
<td>(0.0155)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>6. Unemployed previous year</td>
<td>-0.0042</td>
<td>-0.0277</td>
</tr>
<tr>
<td></td>
<td>(0.0153)</td>
<td>(0.0151)</td>
</tr>
<tr>
<td>7. Poor health previous year</td>
<td>-0.0204</td>
<td>-0.0250</td>
</tr>
<tr>
<td></td>
<td>(0.0221)</td>
<td>(0.0215)</td>
</tr>
<tr>
<td>8. Self-employment</td>
<td>-0.2190</td>
<td>-0.2670</td>
</tr>
<tr>
<td></td>
<td>(0.0297)</td>
<td>(0.0263)</td>
</tr>
<tr>
<td>9. South</td>
<td>-0.1569</td>
<td>-0.0324</td>
</tr>
<tr>
<td></td>
<td>(0.0656)</td>
<td>(0.0333)</td>
</tr>
<tr>
<td>10. Rural</td>
<td>-0.0101</td>
<td>-0.1215</td>
</tr>
<tr>
<td></td>
<td>(0.0317)</td>
<td>(0.0237)</td>
</tr>
<tr>
<td>11. Constant</td>
<td>—</td>
<td>0.8499</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.0433)</td>
</tr>
<tr>
<td>$s^2$</td>
<td>0.0567</td>
<td>0.0694</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>3,135</td>
<td>3,763</td>
</tr>
</tbody>
</table>
Fact

In the random-effects framework, there are two fundamental assumptions. One is that the unobserved individual effects $\alpha_i$ are random draws from a common population. The other is that the explanatory variables are strictly exogenous. That is, the error terms are uncorrelated with (or orthogonal to) the past, current, and future values of the regressors:

$$E(\varepsilon_{it} | x_{i1}, \ldots, x_{iK}) = E(\alpha_i | x_{i1}, \ldots, x_{iK}) = E(v_{it} | x_{i1}, \ldots, x_{iK}) = 0$$
What happens when this condition is violated?

\[ \mathbb{E} (\alpha_i \mid x_{i1}, \ldots, x_{iK}) \neq 0 \quad \text{or} \quad \mathbb{E} (\alpha_i x_{i,t}^\prime) \neq 0 \]

1. The Mundlak’s specification (1978)
2. The Hausman test
Linear unobserved effects panel data models

a. The Mundlak’s specification
Mundlak (1978) criticized the random-effects formulation on the grounds that it neglects the correlation that may exist between the effects $\alpha_i$ and the explanatory variables $x_{it}$. There are reasons to believe that in many circumstances $\alpha_i$ and $x_{it}$ are indeed correlated.

The properties of various estimators we have discussed thus far depend on the existence and extent of the relations between the $X$’s and the effects. Therefore, we have to consider the joint distribution of these variables. However, $\alpha_i$ are unobservable.

Mundlak (1978a) suggested that we approximate $E(\alpha_i x_i,t)$ by a linear function.
Let us assume that the individual effects satisfy:

\[ \alpha_i = \bar{x}_i'a + \alpha_i^* \]

with \( a \in \mathbb{R}^K \) and \( E(\alpha_i^*x_{it}') = 0 \). Under this assumption, the model is:

\[ y_{i,t} = \mu + \beta'x_{i,t} + \bar{x}_i'a + \epsilon_{i,t} \]

\[ \epsilon_{i,t} = \alpha_i^* + \nu_{i,t} \]
Assumptions H$_4$  The error term $\varepsilon_{i,t} = \alpha_i^* + \nu_{i,t}$ satisfies,

$\forall i \in [1, N], \forall t \in [1, T]$

- $E(\alpha_i^*) = E(\nu_{i,t}) = 0$
- $E(\alpha_i^* \nu_{i,t}) = 0$
- $E(\alpha_i^* \alpha_j^*) = \begin{cases} \sigma_{\alpha}^2 & i = j \\ 0 & \forall i \neq j \end{cases}$
- $E(\nu_{i,t} \nu_{j,s}) = \begin{cases} \sigma_{\nu}^2 & t = s, i = j \\ 0 & \forall t \neq s, \forall i \neq j \end{cases}$
- $E(\nu_{i,t} x_i') = E(\alpha_i^* x_i') = 0$
The model can be rewritten as follows:

\[ y_i = \tilde{X}_i^* \gamma + \varepsilon_i \quad \forall i = 1, \ldots, N \]

with \( \varepsilon_i = \alpha_i^* e + v_i \), \( \tilde{X}_i^* = (e \bar{x}_i', e, X_i) \) and \( \gamma' = (a, \mu, \beta') \).

\[
E (\varepsilon_i \varepsilon_j') = E \left[ (\alpha_i^* e + v_i) (\alpha_j^* e + v_j)' \right] \\
= \begin{cases} 
\sigma_{\alpha_i^*}^2 ee' + \sigma_v^2 I_T = V^* & i = j \\
0 & i \neq j 
\end{cases}
\]
Utilizing the expression for the inverse of a partitioned matrix, we obtain the GLS of \((\mu, \beta, a)\) as:

\[
\begin{align*}
\hat{\mu}_{GLS}^* &= \bar{y} - \bar{x}' \hat{\beta}_{BE} \\
\hat{a}_{GLS}^* &= \hat{\beta}_{BE} - \hat{\beta}_{LSDV} \\
\hat{\beta}_{GLS}^* &= \Delta \hat{\beta}_{BE} + (I_K - \Delta) \hat{\beta}_{LSDV}
\end{align*}
\]
Then, the GLS estimator $\hat{\beta}_{GLS}^*$ is unbiased and we have:

$$\hat{\beta}_{GLS}^* \xrightarrow{T \to \infty} \hat{\beta}_{LSDV}$$

$$\hat{\beta}_{GLS}^* \xrightarrow{NT \to \infty} \beta$$
Now let us assume that the DGP corresponds to the Mundlak’s model

\[ \alpha_i = \bar{x}_i' a + \alpha^*_i \]

and we apply GLS to the initial model:

\[ y_{i,t} = \mu + \beta' x_{i,t} + \varepsilon_{i,t} \]

\[ \varepsilon_{i,t} = \alpha_i + \nu_{i,t} \]
Linear unobserved effects panel data models

In general, we have:

$$\hat{\beta}_{GLS} = \Delta \hat{\beta}_{BE} + (I_K - \Delta) \hat{\beta}_{LSDV}$$

In this case, we can show that:

$$E\left(\hat{\beta}_{BE}\right) = \beta + a \quad \hat{\beta}_{BE} \xrightarrow{p} \beta + a \quad \text{as} \quad N \to \infty$$

$$E\left(\hat{\beta}_{LSDV}\right) = \beta$$
Theorem

So, if $\alpha_i = \bar{x}_i a + \alpha_i^*$, the GLS is biased when $T$ is fixed. More precisely:

$$\mathbb{E} \left( \hat{\beta}_{GLS} \right) = \beta + \Delta a$$

$$\hat{\beta}_{GLS} \xrightarrow{p} \beta + \Delta a$$

As usual, the GLS is asymptotically unbiased:

$$\hat{\beta}_{GLS} \xrightarrow{T \to \infty} \beta$$
When $T$ is fixed and $N$ tends to infinity, the GLS is biased if there is a correlation between the individual effects and the explanatory variables:

$$\text{plim}_{N \to \infty} \hat{\beta}_{GLS} = \Delta \text{plim}_{N \to \infty} \hat{\beta}_{BE} + \left( I_K - \Delta \right) \text{plim}_{N \to \infty} \hat{\beta}_{LSDV}$$

$$= \Delta (\beta + a) + \left( I_K - \Delta \right) \beta$$

$$= \beta + \Delta a$$

with $\Delta = \text{plim}_{N \to \infty} \Delta$. 
### Linear unobserved effects panel data models

#### Summary

<table>
<thead>
<tr>
<th>Condition</th>
<th>LSDV</th>
<th>GLS</th>
<th>LSDV</th>
<th>GLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$ Fixed, $N \to \infty$</td>
<td>Unbiased</td>
<td>—</td>
<td>Unbiased</td>
<td>Biased</td>
</tr>
<tr>
<td>$T \to \infty$ and $N \to \infty$</td>
<td>Unbiased</td>
<td>BLUE</td>
<td>Unbiased</td>
<td>Unbiased</td>
</tr>
</tbody>
</table>

$$
E(\alpha_i | x_{i1}, \ldots, x_{iK}) = 0 \quad E(\alpha_i | x_{i1}, \ldots, x_{iK}) \neq 0
$$
b. The Hausman’s specification test
Hausman (1978) proposes a general test of specification, that can be applied in the specific context of linear panel models to the issue of specification of individual effects (fixed or random).

The general idea of the an Hausman’s test is the following. Let us consider a particular model $y = f(x; \beta) + \varepsilon$ and particular hypothesis $H_0$ on this model (parameter, error term etc.). Let us consider two estimators of the $K$-vector $\beta$, denoted $\hat{\beta}_1$ and $\hat{\beta}_2$, both consistent under $H_0$ and asymptotically normally distributed.

1. Under $H_0$, the estimator $\hat{\beta}_1$ attains the asymptotic Cramer–Rao bound.

2. Under $H_1$, the estimator $\hat{\beta}_2$ is biased.
By examining the "distance" between $\hat{\beta}_1$ and $\hat{\beta}_2$, it is possible to conclude about $H_0$:

1. If the "distance" is small, $H_0$ cannot be rejected.
2. If the "distance" is large, $H_0$ can be rejected.
This distance is naturally defined as follows:

$$H = \left( \hat{\beta}_2 - \hat{\beta}_1 \right)' \left[ \text{Var} \left( \hat{\beta}_2 - \hat{\beta}_1 \right) \right]^{-1} \left( \hat{\beta}_2 - \hat{\beta}_1 \right)$$

However, the issue is to compute the variance-covariance matrix $\text{Var} \left( \hat{\beta}_2 - \hat{\beta}_1 \right)$ of the difference between both estimators. Generally we know $\text{V} \left( \hat{\beta}_2 \right)$ and $\text{V} \left( \hat{\beta}_1 \right)$, but not $\text{Var} \left( \hat{\beta}_2 - \hat{\beta}_1 \right)$. 
Lemma (Hausman, 1978)

Based on a sample of $N$ observations, consider two estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ that are both consistent and asymptotically normally distributed, with $\hat{\beta}_1$ attaining the asymptotic Cramer–Rao bound so that $\sqrt{N} (\hat{\beta}_1 - \beta)$ is asymptotically normally distributed with variance–covariance matrix $V_1$. Suppose $\sqrt{N} (\hat{\beta}_2 - \beta)$ is asymptotically normally distributed, with mean zero and variance–covariance matrix $V_2$. Let $\hat{q} = \hat{\beta}_2 - \hat{\beta}_1$. Then the limiting distribution [under the null] of $\sqrt{N} (\hat{\beta}_1 - \beta)$ and $\sqrt{N}\hat{q}$ has zero covariance:

$$E(\hat{\beta}_1 \hat{q}') = 0_K$$
Linear unobserved effects panel data models

Theorem

From this lemma, it follows that

$$\text{Var} \left( \hat{\beta}_2 - \hat{\beta}_1 \right) = \text{Var} \left( \hat{\beta}_2 \right) - \text{Var} \left( \hat{\beta}_1 \right)$$

Thus, Hausman suggests using the statistic

$$H = \left( \hat{\beta}_2 - \hat{\beta}_1 \right)' \left[ \text{Var} \left( \hat{\beta}_2 \right) - \text{Var} \left( \hat{\beta}_1 \right) \right]^{-1} \left( \hat{\beta}_2 - \hat{\beta}_1 \right)$$

or equivalently

$$H = \hat{q}' \left[ \text{Var} \left( \hat{q} \right) \right]^{-1} \hat{q}$$
Under the null hypothesis, this statistic is distributed asymptotically as central chi-square, with $K$ degrees of freedom.

$$ H \xrightarrow{H_0} \chi^2(K) $$

Under the alternative, it has a noncentral chi-square distribution with noncentrality parameter $\bar{q}' [\text{Var}(\hat{q})]^{-1} \bar{q}$, where $\bar{q}$ is defined as follows:

$$ \bar{q} = \lim_{H_1/N \to \infty} \left( \hat{\beta}_2 - \hat{\beta}_1 \right) $$
Problem

Now, apply the Hausman’s test to discriminate between fixed effects methods and random effects methods. We assume that $\alpha_i$ are random variable and the key assumption tested is here defined as:

$$H_0 : E(\alpha_i | X_i) = 0$$

$$H_1 : E(\alpha_i | X_i) \neq 0$$
This test can be interpreted as a specification test between "fixed effect methods" and "random effect methods".

1. If the null is rejected, the correlation between individual effects and the explicative variables induces a bias in the GLS estimates. So, a standard LSDV approach (fixed effect method) has to be privileged.

2. If the null is not rejected, we can use a GLS estimator (random effect method) and specify the individual effects as random variables (random effects model).
How to implement this test? Let us consider the standard model with random effects \((\mu = 0)\):

\[
y_i = X_i \beta + e_\alpha_i + v_i
\]

1. Under \(H_0\) (and assumptions \(H_2\)) we know that \(\hat{\beta}_{LSDV}\) and \(\hat{\beta}_{GLS}\) are consistent and asymptotically normally distributed.

2. Under \(H_0\), \(\hat{\beta}_{GLS}\) is BLUE and attains asymptotic Cramer–Rao bound.

3. Under \(H_1\), \(\hat{\beta}_{GLS}\) is biased.
According to the Hausman’s lemma, we have:

\[ \text{cov} \left( \hat{\beta}_{GLS}, \left( \hat{\beta}_{LSDV} - \hat{\beta}_{GLS} \right) \right) = 0 \iff \text{cov} \left( \hat{\beta}_{LSDV}, \hat{\beta}_{GLS} \right) = \text{Var} \left( \hat{\beta}_{GLS} \right) \]

Since,

\[ \text{Var} \left( \hat{\beta}_{LSDV} - \hat{\beta}_{GLS} \right) = \text{Var} \left( \hat{\beta}_{LSDV} \right) + \text{Var} \left( \hat{\beta}_{GLS} \right) - 2 \text{cov} \left( \hat{\beta}_{LSDV}, \hat{\beta}_{GLS} \right) \]

We have:

\[ \text{Var} \left( \hat{\beta}_{LSDV} - \hat{\beta}_{GLS} \right) = \text{Var} \left( \hat{\beta}_{LSDV} \right) - \text{Var} \left( \hat{\beta}_{GLS} \right) \]
Definition

The Hausman’s specification test statistic of individual effect can be defined as follows:

\[ H = \left( \hat{\beta}_{LSDV} - \hat{\beta}_{GLS} \right) \left[ Var \left( \hat{\beta}_{LSDV} \right) - Var \left( \hat{\beta}_{GLS} \right) \right]^{-1} \left( \hat{\beta}_{LSDV} - \hat{\beta}_{GLS} \right) \]

Under \( H_0 \), we have:

\[ H \xrightarrow{H_0} \chi^2 (K) \]
When $N$ is fixed and $T$ tends to infinity, $\hat{\beta}_{GLS}$ and $\hat{\beta}_{MCG}$ become identical. However, it was shown by Ahn and Moon (2001) that the numerator and denominator of $H$ approach zero at the same speed. Therefore the ratio remains chi-square distributed. However, in this situation the fixed-effects and random-effects models become indistinguishable for all practical purposes.

The more typical case in practice is that $N$ is large relative to $T$, so that differences between the two estimators or two approaches are important problems.
Linear unobserved effects panel data models

Example

Let us consider a simple panel regression model for the total number of strikes days in OECD countries. We have a balanced panel data set for 17 countries ($N = 17$) and annual data form 1951 to 1985 ($T = 35$).

$$s_{i,t} = \alpha_i + \beta_i u_{i,t} + \gamma_i p_{i,t} + \epsilon_{i,t} \quad \forall i = 1, \ldots, 17$$

- $s_{i,t}$ the number of strike days for 1000 workers for the country $i$ at time $t$.
- $u_{i,t}$ the unemployment rate
- $p_{i,t}$, the inflation rate
### Linear unobserved effects panel data models

**PANEL DATA ESTIMATION**

---

**Balanced data:** NI = 17, T = 35, NOB = 595

**WITHIN (fixed effects) Estimates**

**Dependent variable:** SRT

- Sum of squared residuals = 1.46938E+09
- R-squared = 0.242875
- Variance of residuals = 255136
- Adjusted R-squared = 0.219215
- Std. error of regression = 503.110

<table>
<thead>
<tr>
<th>Estimated</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>9.19158</td>
<td>-2.34963</td>
</tr>
<tr>
<td>P</td>
<td>4.75638</td>
<td>3.42113</td>
</tr>
</tbody>
</table>

**Variance Components (random effects) Estimates**

- \(V_{WITH} (\text{variance of } U) = 0.25514E+06\)
- \(V_{BET} (\text{variance of } A) = 55401\)
- (computed from small sample formula)
- \(\Theta = 0.11623\)

**Dependent variable:** SRT

- Sum of squared residuals = 1.52560E+09
- R-squared = 0.214013
- Variance of residuals = 254862
- Adjusted R-squared = 0.139430
- Std. error of regression = 514.647

<table>
<thead>
<tr>
<th>Estimated</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>8.78623</td>
<td>-1.37504</td>
</tr>
<tr>
<td>P</td>
<td>4.76670</td>
<td>3.47489</td>
</tr>
<tr>
<td>C</td>
<td>719.106</td>
<td>5.43737</td>
</tr>
</tbody>
</table>

** Hausman test of HO: FE vs. RE:** CHISQ(Q) = 13.924, P-value = [0.0008]
Section 3.
Random coefficients models
Random coefficients models

There are cases in which there are changing economic structures or different socioeconomic and demographic background factors that imply that the response parameters may be varying over time and/or may be different for different crosssectional units.
3.1. Variable coefficient model
Random coefficient model

When data do not support the hypothesis of coefficients being the same, a single-equation model in its most general form can be written as:

\[ y_{it} = \sum_{k=1}^{K} \beta_{kit} x_{kit} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T \]

where, in contrast to previous sections, we no longer treat the intercept differently than other explanatory variables and let \( x_{1it} = 1 \).
Random coefficient model

Fact

We assume that parameters do not vary with time (the time stability issue is not specific to panel data econometrics)

\[
\beta_{kit} = \beta_{ki}
\]

\[
y_{it} = \sum_{k=1}^{K} \beta_{ki} x_{kit} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T
\]
Random coefficient model

In the case in which only cross-sectional differences are present, this approach is equivalent to postulating a separate regression for each cross-sectional unit

\[ y_i = X_i \beta_i + v_i \quad i = 1, \ldots, N \]

or

\[ y_{it} = \beta_i' x_{it} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T \]

where \( \beta_i = (\beta_{1i}, \beta_{2i}, \ldots, \beta_{Ki}) \) is a \( 1 \times K \) vector of parameters, and \( x_{it} = (x_{1it}, \ldots, x_{Kit}) \) is a \( 1 \times K \) vector of exogenous variables.
Random coefficient model

Definition

Alternatively, each regression coefficient can be viewed as a random variable with a probability distribution (e.g., Hurwicz (1950); Klein (1953); Theil and Mennes (1959); Zellner (1966)).
Random coefficient model

1. The random-coefficient specification reduces the number of parameters to be estimated substantially, while still allowing the coefficients to differ from unit to unit and/or from time to time.

2. Depending on the type of assumption about the parameter variation, it can be further classified into one of two categories: stationary and nonstationary random-coefficient models.
Random coefficient model

**Definition**

Stationary random-coefficient models regard the coefficients as having constant means and variance–covariances. Namely, the $K \times 1$ vector of parameters $\beta_i$ is specified as:

$$\beta_i = \beta + \zeta_i$$

for $i = 1, \ldots, N$, where $\beta$ is a $K \times 1$ vector of constants, and $\zeta_i$ is a $K \times 1$ vector of stationary random variables with zero means and constant variance–covariances.
Random coefficient model

For this type of model we are interested in

1. Estimating the mean coefficient vector $\beta$,
2. Predicting each individual component $\beta_i$,
3. Estimating the dispersion of the individual-parameter vector $\beta_i$ (variance covariance matrix $\Delta$)
4. Testing the hypothesis that the variances of $\xi_i$ (and so $\beta_i$) are zero.
Because of the computational complexities, variable-coefficient models have not gained as wide acceptance in empirical work as has the variable-intercept model. However, that does not mean that there is less need for taking account of parameter heterogeneity in pooling the data.
**Random coefficient model**

**Definition**

When regression coefficients are viewed as invariant over time, but varying from one unit to another, we can write the model as

$$y_{it} = \sum_{k=1}^{K} (\beta_k + \zeta_{ki}) x_{kit} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$ can be viewed as the common mean coefficient $1 \times K$ vector, and $\zeta_i = (\zeta_{1i}, \zeta_{2i}, \ldots, \zeta_{Ki})$ as the individual deviation from the common mean.
Random coefficient model

1. If individual observations are heterogeneous or the performance of individual units from the data base is of interest, then $\zeta_i$ are treated as fixed constants.

2. If conditional on $x_{kit}$, individual units can be viewed as random draws from a common population or the population characteristics are of interest, then $\zeta_i$ are generally treated as random variables having zero means and constant variances and covariances.
Random coefficient model

**Definition (Fixed-coefficient model)**

When $\beta_i$ are treated as fixed and different constants, we can stack the $NT$ observations in the form of the Zellner (1962) seemingly unrelated regression model

$$
\begin{pmatrix}
  y_1 \\
  . \\
  y_N
\end{pmatrix} = 
\begin{pmatrix}
  X_1 & 0 & 0 \\
  0 & .. & .. \\
  0 & .. & X_N
\end{pmatrix}
\begin{pmatrix}
  \beta_1 \\
  . \\
  \beta_N
\end{pmatrix} + 
\begin{pmatrix}
  \nu_1 \\
  . \\
  \nu_N
\end{pmatrix}
$$

where $y_i$ and $\nu_i$ are $T \times 1$ vectors $(y_{it}, ..., y_{iT})$ and $(\nu_{it}, ..., \nu_{iT})$, and $X_i$ is the $T \times K$ matrix of the time-series observations of the $i^{th}$ individual’s explanatory variables with the $t^{th}$ row equal to $x_{it}$. 
Random coefficient model

1. If the covariances between different cross-sectional units are not zero, e.g. $E(v_i v_j) \neq 0$, the GLS estimator of $(\beta'_1, \ldots, \beta'_N)$ is more efficient than the single-equation estimator of $i$ for each cross-sectional unit.

2. If $X_i$ are identical for all $i$ or $E(v_i v_{i'}) = \sigma^2_i I_T$ and $E(v_i v_{j'}) = 0$ for $i \neq j$, the GLS estimator for $(\beta'_1, \ldots, \beta'_N)$ is the same as applying least squares separately to the time-series observations of each cross-sectional unit.
3.2. Random coefficient model
Random coefficient model

**Definition**

We now assume that all the coefficients $\beta_i$ are random, with common mean $\beta$.

$$y_{it} = \sum_{k=1}^{K} (\beta_k + \xi_{ki}) x_{kit} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_K)$ can be viewed as the common mean coefficient $1 \times K$ vector, and $\xi_i = (\xi_{1i}, \xi_{2i}, \ldots, \xi_{Ki})$ as the individual deviation from the common mean. Let us denote

$$\beta_{ki} = \beta_k + \xi_{ki}$$
The vector $x_i = (x_1i...x_{Ki})$ includes a constant term. The parameter $\beta_{ki}$ associated to this constant term correspond to an individual effect.

An alternative notation is:

$$y_{it} = \alpha + \sum_{k=2}^{K} (\beta_k + \zeta_{ki}) x_{kit} + \alpha_i + v_{it}$$

$$E(\alpha_i) = 0$$
Random coefficient model

We will consider a set of assumptions used in the seminal paper of Swamy (1970)

Random coefficient model

Assumptions (Swamy, 1970) Let us assume that

- \( E(\zeta_i) = 0, \ E(v_i) = 0 \)
- \( E(\zeta_i \zeta_j') = \begin{cases} \Delta & i = j \\ 0 & \forall i \neq j \end{cases} \)
- \( E(x_{it} \zeta_j') = 0, \ E(\zeta_i v_j') = 0, \ \forall (i, j) \)
- \( E(v_i v_j') = \begin{cases} \sigma_i^2 I_T & t = s, i = j \\ 0 & \forall t \neq s, \forall i \neq j \end{cases} \)
Remark: we assume that the error term $v_i$ is heteroskedastic:

$$E(v_i v_i') = \sigma_i^2 I_T$$
Random coefficient model

Definition

The two first moments of the vector of random parameters \( \beta_i = \beta + \xi_i \) are defined by, \( \forall i = 1, \ldots, N \):

\[
\begin{align*}
E(\beta_i) &= \beta \\
\Delta &= E(\beta_i \beta_i') = E(\xi_i \xi_i') = \\
&= \begin{pmatrix}
\sigma_{\xi_1}^2 & \sigma_{\xi_1,\xi_2} & \cdots & \sigma_{\xi_1,\xi_K} \\
\sigma_{\xi_2,\xi_1} & \sigma_{\xi_2}^2 & \cdots & \sigma_{\xi_2,\xi_K} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{\xi_K,\xi_1} & \sigma_{\xi_K,\xi_2} & \cdots & \sigma_{\xi_K}^2
\end{pmatrix}
\end{align*}
\]
Random coefficient model

Fact

The matrix $\Delta$ corresponds to variance covariance matrix of the random parameters $\beta_i = (\beta_1, \beta_2, \ldots, \beta_K)'$, which is assumed to be common to all cross section units:

$$
\Delta_{(K,K)} = 
\begin{pmatrix}
\sigma_{\beta_1}^2 & \sigma_{\beta_1,\beta_2} & \ldots & \sigma_{\beta_1,\beta_K} \\
\sigma_{\beta_2,\beta_1} & \sigma_{\beta_2}^2 & \ldots & \sigma_{\beta_2,\beta_K} \\
\ldots & \ldots & \ldots & \ldots \\
\sigma_{\beta_K,\beta_1} & \sigma_{\beta_K,\beta_2} & \ldots & \sigma_{\beta_K}^2 
\end{pmatrix}
$$
Random coefficient model

Definition

For each cross section unit, we have:

\[ y_i = X_i \beta + X_i \xi_i + v_i \]

with

\[ \beta_i = \beta + \xi_i \]

where the vector \( X_i \) include a constant term (i.e. the average of random individual effects, \( \alpha \)).
Random coefficient model

Definition

This model can be rewritten as follows:

\[ y_i = X_i \beta + \varepsilon_i \]

with

\[ \varepsilon_i = X_i \xi_i + v_i = X_i (\beta_i - \beta) + v_i \]
Random coefficient model

For a given cross unit, the covariance matrix for the composite disturbance term $\varepsilon_i = X_i\tilde{\zeta}_i + v_i$ is defined by:

$$\Phi_i = E(\varepsilon_i\varepsilon_i')$$

$$= E[(X_i\tilde{\zeta}_i + v_i)(X_i\tilde{\zeta}_i + v_i)']$$

$$= X_iE(\tilde{\zeta}_i\tilde{\zeta}_i')X_i' + E(v_iv_i')$$
Definition

For a given cross unit, the covariance matrix for the composite disturbance term $\varepsilon_i = X_i \zeta_i + v_i$ is defined by:

$$\Phi_i = X_i \Delta X_i' + \sigma_i^2 I_T$$

Stacking all $NT$ observations, the covariance matrix for the composite disturbance term is either block-diagonal and heteroskedastic or diagonal and heteroskedastic.
Under Swamy’s assumption, the simple regression of $y$ on $X$ will yield an unbiased and consistent estimator of $\beta$ if $(1/NT)X'X$ converges to a nonzero constant matrix.

But the estimator is inefficient, and the usual least-squares formula for computing the variance–covariance matrix of the estimator is incorrect, often leading to misleading statistical inferences.
Random coefficient model

Definition

The best linear unbiased estimator of $\beta$ is the GLS estimator

$$\hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} X_i' \Phi_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' \Phi_i^{-1} y_i \right)$$
Random coefficient model

**Definition**

The GLS estimator $\hat{\beta}_{GLS}$ is a matrix-weighted average of the least-squares estimator $\hat{\beta}_i$ for each cross-sectional unit, with the weights inversely proportional to their covariance matrices:

$$\hat{\beta}_{GLS} = \sum_{i=1}^{N} W_i \hat{\beta}_i$$

$$W_i = \left\{ \sum_{i=1}^{N} \left[ \Delta + \sigma_i^2 (X'_i X_i)^{-1} \right]^{-1} \right\}^{-1} \left[ \Delta + \sigma_i^2 (X'_i X_i)^{-1} \right]^{-1}$$

$$\hat{\beta}_i = (X'_i X_i)^{-1} X'_i y_i$$
Random coefficient model

The covariance matrix for the GLS estimator is:

\[
V \left( \hat{\beta}_{GLS} \right) = \left( \sum_{i=1}^{N} X_i' \Phi_i^{-1} X_i \right)^{-1}
\]

\[
= \left\{ \sum_{i=1}^{N} \left[ \Delta + \sigma_i^2 (X_i' X_i)^{-1} \right]^{-1} \right\}^{-1}
\]
Random coefficient model

Swamy proposed using the least-squares estimators
\[ \hat{\beta}_i = (X_i'X_i)^{-1} X_i' y_i \]
and their residuals \[ \hat{v}_i = y_i - X_i \hat{\beta}_i \]
to obtain unbiased estimators of \( \sigma_v^2 \) and \( \Delta \)

\[ \hat{\sigma}_i^2 = \frac{1}{T - K} y_i' \left[ I - X_i (X_i'X_i)^{-1} X_i' \right] y_i \]
\[ = \frac{1}{T - K} \sum_{t=1}^{T} \hat{v}_{i,t} \]

with \( y_{i,t} = \hat{\beta}_i x_{i,t} + \hat{v}_{i,t} \).
Random coefficient model

For the $\Delta$ matrix, we have:

$$
\hat{\Delta}_{(K,K)} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right) \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right)' \right] \\
- \frac{1}{N} \sum_{i=1}^{N} \tilde{\sigma}_i^2 (X'_i X_i)^{-1}
$$
Random coefficient model

Definition

However, the previous estimator $\hat{\Delta}$ is not necessarily nonnegative definite. In this situation, Swamy (1970) has suggested replacing this estimator by:

$$\hat{\Delta}_{(K,K)} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right) \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right)' \right]$$

This estimator, although not unbiased, is nonnegative definite and is consistent when $T$ tends to infinity.
Random coefficient model

Swamy proved that substituting $\hat{\sigma}_i^2$ and $\hat{\Delta}$ for $\sigma_i^2$ and $\Delta$ yields an asymptotically normal and efficient estimator of $\beta$. The speed of convergence of the GLS estimator is $N^{1/2}$. 
Random coefficient model

Summary: how to estimate a random coefficient model?

1. Run the $N$ individual regressions $y_{i,t} = \hat{\beta}_i' x_{i,t} + \hat{\nu}_{i,t}$.

2. Compute $\hat{\sigma}_i^2$ and the Swamy’s estimator $\hat{\Delta}$ as follows

\[
\hat{\Delta}_{(K,K)} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right) \left( \hat{\beta}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \right)' \right]
\]

\[
\hat{\sigma}_i^2 = \frac{1}{T-K} y_i' \left[ I - X_i (X_i' X_i)^{-1} X_i' \right] y_i
\]
3. Compute the GLS estimate of the expectation of individual parameters $\beta_i$

$$ \hat{\beta}_{GLS} = \left( \sum_{i=1}^{N} X'_i \hat{\Phi}_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^{N} X'_i \hat{\Phi}_i^{-1} y_i \right) $$

$$ \hat{\Phi}_i = X_i \Delta X'_i + \hat{\sigma}_i^2 I_T $$
Random coefficient model

Example

Swamy (1970) used this model to reestimate the Grunfeld investment function with the annual data of 11 U.S. corporations. His GLS estimates of the common-mean coefficients of the firms’ beginning-of-year value of outstanding shares and capital stock are 0.0843 and 0.1961, with asymptotic standard errors 0.014 and 0.0412, respectively. The estimated dispersion measure of these coefficients is

$$\hat{\Delta} = \begin{pmatrix} 0.0011 & -0.0002 \\ 0.0187 & 0.0187 \end{pmatrix}$$
Predicting Individual Coefficients

1. Sometimes one may wish to predict the individual component $\beta_i$, because it provides information on the behavior of each individual and also because it provides a basis for predicting future values of the dependent variable for a given individual.

2. Swamy (1970, 1971) has shown that the best linear unbiased predictor, conditional on given $\beta_i$, is the least-squares estimator $\hat{\beta}_i$. 
Random coefficient model

Definition

Lee and Griffiths (1979) suggest predicting $\beta_i$ by

\[
\beta_i^* = \hat{\beta}_{GLS} + \Delta X_i' \left( X_i \Delta X_i' + \sigma_i^2 I_T \right)^{-1} \left( y_i - X_i \hat{\beta}_{GLS} \right)
\]

This predictor is the best linear unbiased estimator in the sense that $E \left( \beta_i^* - \beta_i \right) = 0$, where the expectation is an unconditional one.
3.3. Mixed fixed and random coefficient model
Many of the previously discussed models can be treated as special cases of a general mixed fixed- and random-coefficients model.
We assume that each cross section unit is different

\[ y_{it} = \sum_{k=1}^{K} \beta_{ki} x_{ki} + \sum_{l=1}^{m} \gamma_{li} w_{li} + v_{it} \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T \]

where \( x_{it} \) and \( w_{it} \) are each a \( K \times 1 \) and an \( m \times 1 \) vector of explanatory variables that are independent of the error of the equation \( v_{it} \).
Random coefficient model

The parameters \( \beta = (\beta'_1, \beta'_2, \ldots, \beta'_K)' \) are assumed to be random and parameters \( \gamma = (\gamma'_1, \gamma'_2, \ldots, \gamma'_m)' \) are fixed.
Random coefficient model

In a vectorial form, we have

\[
y_{(NT,1)} = X_{(NT,NK)} \beta_{(NK,1)} + W_{(NT,Nm)} \gamma_{(Nm,1)} + v_{(NT,1)}
\]

\[
X = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & X_N & 0 \end{pmatrix} \quad W = \begin{pmatrix} W_1 & 0 & 0 \\ 0 & W_2 & 0 \\ 0 & 0 & W_N \end{pmatrix}
\]
Random coefficient model

The motivation of a mixed fixed and random coefficients model is that while there may be fundamental differences among cross-sectional units, by conditioning on these individual specific effects one may still be able to draw inferences on certain population characteristics through the imposition of a priori constraints on the coefficients of $x_{it}$ and $w_{it}$. 
Random coefficient model

We assume that there exist two kinds of restrictions, stochastic and fixed, in the following form:

1. The coefficients of $x_{it}$ are assumed to be subject to stochastic restrictions of the form

$$\beta = A_1 \bar{\beta} + \zeta$$

where $A_1$ is an $NK \times L$ matrix with known elements, $\bar{\beta}$ is an $L \times 1$ vector of constants, and $\zeta$ is assumed to be (normally distributed) random variables with mean 0 and nonsingular constant covariance matrix $C$ and is independent of $x_i$. 

C. Hurlin

Panel Data Econometrics
2. The coefficients of $w_{it}$ are assumed to be subject to

$$\gamma = A_2 \bar{\gamma}$$

where $A_2$ is an $Nm \times n$ matrix with known elements, and $\bar{\gamma}$ is an $n \times 1$ vector of constants. Since $A_2$ is known, we can rewrite the model as

$$y_{(NT,1)} = X_{(NT,NK)} \beta_{(NK,1)} + \tilde{W}_{(NT,n)} \gamma_{(n,1)} + v_{(NT,1)}$$

where $\tilde{W} = WA_2$. 
Random coefficient model

Many of the linear panel data models with unobserved individual specific but time-invariant heterogeneity can be treated as special cases of this model.
Random coefficient model

Example

A common model for all cross-sectional units. If there is no interindividual difference in behavioral patterns, we may let $X = 0$, $A_2 = e_N \otimes I_m$, so model becomes

$$y_{it} = w_{it} \bar{\gamma} + v_{it}$$
Random coefficient model

Example

When each individual is considered different, then $X = 0$, $A_2 = I_N \otimes I_m$, and the model becomes

$$y_{it} = w_{it} \gamma_i + \nu_{it}$$
Example

When the effects of the individual specific, time-invariant omitted variables are treated as random variables just as in the assumption on the effects of other omitted variables, we can let $X_i = e_T$, $\zeta' = (\zeta_1, \ldots, \zeta_N)$, $A_1 = e_N$, $C = I_N \sigma^2_{\alpha}$, $\bar{\beta}$ be an unknown constant, and $w_{it}$ not contain an intercept term. Then the model becomes:

$$y_{it} = \bar{\beta} + \bar{\gamma}' w_{it} + \zeta_i + v_{it}$$