Panel Data: Fixed Effect Models

Thierry Magnac

October 2008

1 Introduction

Panel data of households, countries, firms ... over several time periods and is more elaborate than the pooling of cross sections (see Pseudo-Panels).

Examples:

- PSID : Income, Food consumption
- NLS: Labor, income
- HRS: Older people, retirement; health
- ECHP: Labor, income ...
- German SEP
- World Bank surveys
- Matched Employer/employee data
- Consumer panels

References:

- Arellano, M., 2003, *Panel Data Econometrics*, Cambridge UP, Cambridge.
- Hsiao, C., 2004, *Panel Data*, Cambridge UP.

Gains*:*

- More information, more variability
- Better ability to identify effects
  Example 1: Age and cohort (but not time, age and cohort)
  Example 2: 50% of women are found to be working at t. Does it mean that:
  (a) Half of them work all the time.
  (b) All of them work half time.
- Controlling for heterogeneity

  Example: $y_{it} = x_{it}\beta + \gamma D_{it} + \varepsilon_{it}$
  where $D_{it} =$ "treatment" dummy variable (1 or 0)
  Unionization, training last period, regulating the market etc

  The issue is the correlation between D_{it} and ε_{it} i.e. state dependence cannot be distinguished from unobserved heterogeneity
  What panel data brings up is the possibility of differencing an individual effect from which could come the endogeneity bias.
- Dynamics

but some limitations:

- Costly design and data collection
- Measurement errors could be more severe

- Selectivity: non response and attrition.

2 Fixed Effects Models

Panel data; double index y_{it} and x_{it} for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
Hypothesis: no missing value so the sample is balanced, y_{it} and x_{it} are observed for any i and t.

2.1 Definition

The model is:

$$\begin{aligned} y_{it} &= x_{it}\beta + \varepsilon_{it} \\ &= x_{it}\beta + \alpha_i + \delta_t + u_{it} \end{aligned}$$

by the variance decomposition of ε_{it}.

Parameters: δ_t are macro shocks, α_i are individual effects, the parameter of interest is β.

2.1.1 Period effects and time dummies

Hypothesis: For simplicity, $\{\delta_t\}_{t=1,.,T}$ are included among the x_{it} as coefficients of period indicators or time dummies $x_{it}^{(\tau)}$ such that $x_{it}^{(\tau)} = \mathbf{1}\{t = \tau\}$.

Rule: Always include time dummies at least in the first run of the model so that you can test for their presence. Always justify their absence.

In contrast, the α_i have almost no interest in most cases so that they are considered as nuisance parameters.
We have:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}.$$

2.1.2 Assumptions

Let $x_i = (x_{i1}, ., x_{iT})$ a matrix of dimension $[T, K]$;

The *principal or strict exogeneity assumption* is

$$H_1 : E(x_i' u_{it}) = 0, E(u_{it}) = 0$$

the second part of H_1 enables the identification of α_i by centering.
This assumption could be replaced by a stronger mean independence assumption $E(u_{it} \mid x_i) = 0$.

Specific assumptions (u_i is a column vector, $u_i = (u_{i1}, ., u_{iT})$.

$$H_2^a : u_i \text{ and } u_j \text{ are independent}$$

$$H_2^b : E(u_i u_i' \mid x_i) = \sigma_u^2 . I_T$$

We never question the first part H_2^a. Homoskedasticity H_2^b is for simplicity.

Conclusion: Under H_1 and H_2, the fixed effect model seems to be estimable by OLS.

Two worries:

- Identification
- How to get rid of the nuisance parameters whose estimates cannot be consistent if $T \nrightarrow \infty$.

2.2 Example: Structural economic model

Intertemporal labor supply:

$$\begin{aligned} \max \quad & \sum_{t=0}^{T} \beta^t U_{it}(l_{it}, c_{it}) \\ s.t. \quad & \sum_{t=0}^{T} \frac{1}{(1+r)^t}(p_t c_{it} + w_{it} l_{it}) = A_{0i} + \sum_{t=0}^{T} \frac{w_{it}}{(1+r)^t} \end{aligned}$$

Specification assumption:

$$U_{it}(l_{it}, c_{it}) = -\alpha_{it}(1 - l_{it})^\gamma + v(c_{it})$$

To make sure that $\frac{\partial U_{it}}{\partial l_{it}} > 0, \frac{\partial^2 U_{it}}{\partial l_{it}^2} < 0$, we shall impose $\gamma > 1, \alpha_{it} > 0$.
The first order condition is written:

$$\beta^t \frac{\partial U_{it}}{\partial l_{it}} = \frac{w_{it}}{(1+r)^t}.\lambda_i$$

where λ_i is the Lagrange multiplier of the intertemporal budget constraint or the marginal value of assets.
Replacing $l_{it} = 1 - h_{it}$ we get:

$$\alpha_{it}\gamma\beta^t h_{it}^{\gamma-1} = \frac{w_{it}}{(1+r)^t}.\lambda_i$$

thus:

$$\log h_{it} = \frac{1}{\gamma - 1} \log w_{it} + \delta_t + \tilde{\alpha}_i + \varepsilon_{it}.$$

where $\frac{1}{\gamma-1}$ is the intertemporal elasticity of substitution, δ_t is a function of the preferences for the future and $\tilde{\alpha}_i$ mixes preferences for leisure and the marginal utility of wealth. ε_{it} is a preference shock.

3 Identification and Geometric Interpretation.

Example: If one of the x_{it} is the intercept we can write:

$$\begin{aligned} y_{it} &= x_{it}^{(-0)}\beta_1 + \beta_0 + \alpha_i + u_{it} \\ &= x_{it}^{(-0)}\beta_1 + \alpha_i^{(0)} + u_{it} \end{aligned}$$

where $x_{it}^{(-0)}$ means explanatory variables without the constant and where $\alpha_i^{(0)} = \alpha_i + \beta_0$.
We say that $(0, \alpha_i^{(0)})$ is *observationally equivalent* to (β_0, α_i).

3.1 Identification

Split explanatory variables into time-varying and non time-varying variables, $x_{it} = (x_{it}^{(1)}, x_i^{(2)})$. Then:

$$\begin{aligned} y_{it} &= x_{it}^{(1)}\beta_1 + x_i^{(2)}\beta_2 + \alpha_i + u_{it} \\ &= x_{it}^{(1)}\beta_1 + \alpha_i^{(2)} + u_{it} \end{aligned}$$

where $\alpha_i^{(2)} = x_i^{(2)}\beta_2 + \alpha_i$. Thus, $(\beta_1, 0, \alpha_i^{(2)})$ is observationally equivalent to $(\beta_1, \beta_2, \alpha_i)$.

Time varying? The definition is ad-hoc so that the *parameter of interest* β_1 is identified. Write:

$$\begin{aligned} y_{it} &= x_{it}^{(1)}\beta_1 + \alpha_i^{(2)} + u_{it} \\ &= (x_{it}^{(1)} - x_{i.}^{(1)})\beta_1 + \tilde{\alpha}_i + u_{it} \end{aligned}$$

where $x_{i.}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} x_{it}^{(1)}$ the individual mean of these variables.
Assume that:

$$rank(E((x_i^{(1)} - x_{i.}^{(1)})'(x_i^{(1)} - x_{i.}^{(1)}))) = \dim(\beta_1).$$

where $x_i^{(1)} - x_{i.}^{(1)}$ is a $[T, K]$ matrix of elements $x_{ikt}^{(1)} - x_{ik.}^{(1)}$.

Remark: Additional assumptions needed to identify β_2.

3.2 Projections

Stack the dependent variable of dimension NT into:

$$Y' = (y_{11}...y_{1T}...y_{i1}...y_{iT}...y_{N1}...y_{NT})$$

Denote the individual mean of any variable as:

$$y_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}.$$

Definition 1 : The operator B, called *between operator* transforms any vector Y of elements y_{it} into the vector, BY, of dimension NT, which elements are individual means, $y_{i.}$.

Definition 2 : The operator W called *within operator* transforms any vector Y of elements y_{it} into the vector, WY, of dimension NT, which elements are the differences to individual means, $y_{it} - y_{i.}$.

Notice that:

$$W = I - B$$

where I is the identity matrix in $\mathbb{R}^{NT}$.

By construction, these operators enable the decomposition of any vector of $\mathbb{R}^{NT}$ into its projections in two sub-spaces orthogonal betwen them.

$$I = B + W \quad BW = 0$$

$$B^2 = B = B' \quad W^2 = W = W'$$

$$\left\|\tilde{Y}\right\|^2 = \left\|B\tilde{Y}\right\|^2 + \left\|W\tilde{Y}\right\|^2$$

where $\|.\|^2 = \sum_{i,t} x_{it}^2$ and where $\tilde{Y}$ has elements $y_{it} - y_{..}$.

The matrix expression for the fixed effect model is:

$$Y = X\beta + \sum_{i=1}^{N} \alpha_i D_i + U$$

where elements of vectors D_i are all zero except the ones related to individual i: between locations $i(T-1)+1$ and iT, elements are equal to 1.
These vectors are non stochastic and:

$$BD_i = D_i, \quad WD_i = 0,$$

and vectors D_i form a basis of the inter-individual space.

Hypotheses H_1 and H_2 can be written:

$$E(X'U) = 0 \qquad E(U) = 0$$

$$E(UU' \mid X) = \sigma_u^2 I_{NT}$$

We can thus decompose the fixed effect model into its between and within dimensions:

$$\begin{cases} WY = WX\beta + WU \\ BY = BX\beta + \sum_{i=1}^{N} \alpha_i D_i + BU \end{cases}$$

and by the projection properties, both models are equivalent.

4 Within Estimation

4.1 Definition and Asymptotic Properties

The model projected in the between dimension is not informative on β:

$$BY = \sum_{i=1}^{N}(\alpha_i + x_{i.}\beta)D_i + BU = \sum_{i=1}^{N}\tilde{\alpha}_i D_i + BU$$

The within dimension only is informative:

$$WY = WX\beta + WU$$

Assume that β is identified in the within dimension that is (the time varying condition):

$$H3 : rank(E((x_i^{(1)} - x_{i.}^{(1)})'(x_i^{(1)} - x_{i.}^{(1)})) = K.$$

Since:

$$E(WUU'W) = \sigma^2 W = \sigma^2\Omega,$$

we are in the case where ΩX is proportional to X so that OLS is BLUE in the projected model. Define the within estimator as the OLS estimator in the within dimension:

$$\hat{\beta}_w = (X'WX)^{-1}X'WY.$$

Because of $H3$, matrix $(X'WX)$ is invertible at least when the number of observations is large.

The variance of this estimator is:

$$V\widehat{\beta}^{(w)} = \sigma_u^2 \left(X'WX\right)^{-1}.$$

Since the within estimator is an OLS estimator, it is consistent and asymptotically normal:

$$\sqrt{n}(\hat{\beta}_W - \beta) \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, \sigma_u^2 \left(E((x_i^{(1)} - x_{i.}^{(1)})'(x_i^{(1)} - x_{i.}^{(1)}))\right)^{-1})$$

Remark: When n and T tends to infinity, the estimation remains consistent and asymptotically normal if:

$$\lim_{T\to\infty} E(\frac{1}{T}(x_i^{(1)} - x_{i.}^{(1)})'(x_i^{(1)} - x_{i.}^{(1)})) = V_0$$

is definite positive.

4.2 Estimation of fixed effects

Use the between model

$$B(Y - X\beta) = \sum_{i=1}^{N} \alpha_i D_i + BU$$

which is akin to an analysis of variance except that β is replaced by its estimator $\hat{\beta}_w$. Thus:

$$\hat{\alpha}_i = y_{i.} - x_{i.}\widehat{\beta}^{(w)}$$

This estimation is individual by individual. It is thus consistent only in the time dimension. Their distribution function might be of more interest (See next chapter).

Remark: Testing $\mathrm{H}_0 : \alpha_i = \alpha$ is as in an analysis of variance but not very interesting.

4.3 Orthogonal deviations

There are other possible transformations of the data to remove the fixed effect, for instance first differences:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}.$$

Yet, it introduces serial correlation because for instance $Cov(\Delta u_{it}, \Delta u_{it-1}) = -\sigma^2$ in the case where the original disturbances are homoskedastic. To get rid of this serial correlation we should do GLS and "sphericize' the equation.
Specifically, we define *forward orthogonal deviations* as:

$$u_{it}^* = \frac{T-t}{T-t+1}\left[u_{it} - \frac{1}{T-t}(u_{it+1} + ... + u_{iT})\right] \text{ for } t \leq T-1.$$

where $Vu_{it}^* = \sigma^2 I_{T-1}$.

Estimation will follow the same structure as before and the within group estimator is also the one using orthogonal deviations.

5 Heteroskedasticity & Serial Correlation

Deal with heteroskedasticity of unknown form including serial correlation. Let $u_i = (u_{i1}, ., u_{iT})'$ and

$$E(u_i u_i' \mid x_i) = \Sigma(x_i).$$

5.1 Within estimation

As OLS in the case of heteroskedasticity and serial correlation, the within estimator is consistent but its standard errors are incorrect:

$$\hat{\beta}_w - \beta = (X'WX)^{-1}X'WU$$

Write:

$$\begin{aligned} X'WU &= \sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - x_{i.})'u_{it} \\ &= \sum_{i=1}^{n}(x_i - x_{i.})'u_i. \end{aligned}$$

We know that by the principal assumption:

$$E((x_i - x_{i.})'u_i) = 0,$$

and:

$$\begin{aligned} V((x_i - x_{i.})'u_i) &= E((x_i - x_{i.})'u_i u_i'(x_i - x_{i.})) \\ &= E((x_i - x_{i.})'E(u_i u_i' \mid x_i)(x_i - x_{i.})) \\ &= E((x_i - x_{i.})'\Sigma(x_i)(x_i - x_{i.})) \equiv V_{XWU}. \end{aligned}$$

By the central limit theorem:

$$\sqrt{n}\frac{X'WU}{n} \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, V_{XWU}).$$

and by the LLN:

$$plim_{n\to\infty}\frac{X'WX}{n} = E((x_i - x_{i.})'(x_i - x_{i.})) \equiv V_{WX}.$$

Combining results:

$$\sqrt{n}(\hat{\beta}_w - \beta) \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, \sigma^2(V_{WX})^{-1}.V_{XWU}.(V_{WX})^{-1}).$$

How do we get a consistent estimate of the variance-covariance matrix if $n \to \infty$? Replace by empirical counterparts, including u_i by $\hat{u}_i$, since $\hat{\beta}_w$ is consistent for β i.e.:

$$\begin{aligned}\hat{V}_{WX} &= \frac{X'WX}{n}\\ \hat{V}_{XWU} &= \frac{X'W\hat{U}\hat{U}'WX}{n}.\end{aligned}$$

Remark: When n is fixed and T tends to infinity, the previous estimate is not consistent. Another consistent estimate can be designed that relies on serial dependence decreasing fastly over time. See Newey and West, 1987.

5.2 GLS and GMM

When heteroskasticity or serial correlation is present, the optimality of within estimation is lost. Strengthen the principal assumption into:

$$E(u_i \mid x_i) = 0$$

while $E(u_i u_i' \mid x_i) = \Sigma(x_i)$.

There are two possibilities. An unfeasible GLS estimator is given by:

$$\hat{\beta}_{UGLS} = \left(\frac{1}{n}\sum(x_i - x_{i.})' \left(\Sigma(x_i)\right)^{-1}(x_i - x_{i.})\right)^{-1}\left(\left(\frac{1}{n}\sum(x_i - x_{i.})' \left(\Sigma(x_i)\right)^{-1}(y_i - y_{i.})\right)^{-1}\right)$$

It is unfeasible because $\Sigma(x_i)$ is unknown.

Remark: Given some non parametric estimate, $\hat{\Sigma}(x_i)$, we can derive a FGLS estimate.

Another simpler possibility is to use GMM. The moment conditions that can be used are:

$$E(x_{i\tau}(u_{it} - u_{i.})) = 0 \text{ for any } (\tau, t).$$

and therefore:

$$E(x_{i\tau}(y_{it} - y_{i.} - (x_{it} - x_{i.})\beta)) = 0 \text{ for any } (\tau, t).$$

Write z_i a matrix of dimension $[T^2K, T]$ which stacks $x_{i\tau}$ in order to have the T^2K moment conditions (with some redundancy):

$$E(z_i'(y_i - y_{i.} - (x_i - x_{i.})\beta)) = 0.$$

The GMM estimator is thus:

$$\hat{\beta}_{GMM} = \left[\left(\sum_i (x_i - x_{i.})' z_i\right) S_n \left(\sum_i z_i'(x_i - x_{i.})\right)\right]^{-1} \left(\sum_i (x_i - x_{i.})' z_i\right) S_n \left(\sum_i z_i'(y_i - x_{i.})\right)$$

where the optimal choice of S_n is the estimate of the inverse of the variance of the moment conditions:

$$S_n = \left(\sum_i z_i'(y_i - y_{i.} - (x_i - x_{i.})\hat{\beta}_{GMM})(y_i - y_{i.} - (x_i - x_{i.})\hat{\beta}_{GMM})' z_i \right)^-$$

Remark: In terms of efficiency, the estimators are ranked by (the inverse of) their asymptotic covariances as:

$$\hat{\beta}_{UGLS} \succeq \hat{\beta}_{GMM} \succeq \hat{\beta}_W$$

while they all collapse into the same estimate under the homoskedasticity assumption.

Panel Data: Random Effect Models

Thierry Magnac

October 2007

Idea: As α_i is a nuisance parameter, we treat α_i as a random variable.

1 Model specification

As before, using a different interpretation:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

where α_i is individual unobserved heterogeneity, constant over time, while u_{it} is unobserved heterogeneity which is variable over time. Time dummies are always included.

1.1 Definition

Two types of variables: time varying variables, $x_{it}^{(1)}$, and variables $x_i^{(2)}$:

$$y_{it} = x_{it}^{(1)}\beta_1 + x_i^{(2)}\beta_2 + \alpha_i + u_{it}$$

Suppose that the coefficients of the time-varying variables are identified:

$$H_4 : E((x_i^{(1)} - x_{i.}e_T)'(x_i^{(1)} - x_{i.}e_T)) = \dim(\beta_1)$$

Specification hypothesis: If $x_i = (x_{i1}^{(1)}, ..., x_{iT}^{(1)}, x_i^{(2)})$:

$$H_1 : Ex_i'u_{it} = 0, Eu_{it} = 0$$

$$H_2^a : u_i \text{ and } u_j \text{ are independent}$$

$$H_2^b : E(u_i u_i' \mid x_i) = \sigma_u^2 . I_T$$

$$H_3 : E\alpha_i = 0, E(\alpha_i u_{it} \mid x_i) = 0$$

The structure of the error term is called an *error component* or *random effect* structure.

1.2 Claims

Claim 1: Without any other assumption, this model is equivalent to the previous one.
Claim 2: Simple OLS is biased because the component α_i and the explanatory variables x_i are correlated.

2 Mundlak's model

2.1 Correlated individual effects

Idea: Control the correlation between individual effects and explanatory variables. Write:

$$\alpha_i = x_{i.}^{(1)} \theta_1 + x_i^{(2)} \theta_2 + v_i = z_i \theta + v_i$$

so that $E(z_i' v_i) = 0$. It is a definition of θ (see Proof)

Write:

$$y_{it} = (x_{it}^{(1)} - x_{i.}^{(1)})\beta_1 + x_{i.}^{(1)}\gamma_1 + x_i^{(2)}\gamma_2 + v_i + u_{it}$$

where (see Proof):

$$\gamma_1 = \beta_1 + \theta_1, \qquad \gamma_2 = \beta_2 + \theta_2.$$

There are two consequences:

- Without any other hypothesis about θ_1 and θ_2, there is no dependence between parameters β_1 on the one hand and parameters γ_1 and γ_2 on the other hand.
- The model can be written as:

$$Y = WX_1\beta_1 + BX_1\gamma_1 + BX_2\gamma_2 + \tilde{\varepsilon}$$

where we redefine the error $\tilde{\varepsilon}$ in an appropriate way:

$$\tilde{\varepsilon}_{it} = v_i + u_{it}.$$

2.2 Identity with Fixed Effects

Projecting on within and between dimensions, we get two groups of orthogonal regressors: WX_1 on the one hand, BX_1 and BX_2 on the other.

$$\begin{cases} WY = WX_1\beta_1 + W\tilde{\varepsilon} \\ BY = BX_1\gamma_1 + BX_2\gamma_2 + B\tilde{\varepsilon} \end{cases}$$

Note that:

$$B\tilde{\varepsilon} = \{v_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}\}_{i,t} \quad W\tilde{\varepsilon} = \{u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}\}_{i,t} = W\varepsilon$$

Hypothesis H_1 of absence of correlation between shocks u_{it} and explanatory variables translates into:

$$E(X_1'W\varepsilon) = 0$$

Proposition: *The OLS estimator β_1 in Mundlak's model is the covariance or within estimator:*

$$\hat{\beta}_1^{(M)} = \hat{\beta}_1^{(w)}$$

2.3 The between estimator

Write the model in its individual dimension:

$$y_{i.} = x_{i.}^{(1)}\gamma_1 + x_i^{(2)}\gamma_2 + v_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}$$

Definition: The OLS estimator projected in the between dimension is called the “between” estimator.

Hypothesis H_1 supposes the absence of correlation between u_{it} and explanatory variables and by construction, v_i, is not correlated with explanatory variables ($E(z_i' v_i) = 0$). The between estimator is consistent.

Assume that individual effects, v_i, are homoskedastic and replace H_3 by H_3':

$$H_3'^{(a)} : v_i \text{ and } v_j \text{ are independent}$$

$$H_3'^{(b)} : E(z_i' v_i) = 0, \quad E(v_i u_{it} \mid z_i) = 0, \quad E(v_i^2 \mid z_i) = \sigma_v^2$$

where the first part of $H_3'^{(b)}$ is true by construction of v_i and where the second part of $H_3'^{(b)}$ is an assumption.
The model in its between dimension is homoskedastic:

$$E\left[\left(v_i + \frac{1}{T}\sum_{t=1}^{T} u_{it}\right)^2 \mid z_i\right] = \sigma_v^2 + \frac{\sigma_u^2}{T}$$

so that the between estimator denoted $\hat{\gamma}^{(b)}$, is consistent and asymptotically normal and BLUE in the between dimension.

2.4 Recombining models

Mundlak's model:

$$Y = WX_1\beta_1 + BX_1\gamma_1 + BX_2\gamma_2 + \tilde{\varepsilon}$$

The two groups of regressors (WX_1 and BX_1, BX_2) are orthogonal so that the OLS estimators in the global model is equal to the OLS estimators in the projected dimensions.

The derived model, called the *Mundlak procedure*, is:

$$\begin{aligned} y_{it} &= (x_{it}^{(1)} - x_{i.}^{(1)})\beta_1 + x_{i.}^{(1)}\gamma_1 + x_i^{(2)}\gamma_2 + v_i + u_{it} \\ &= x_{it}^{(1)}\beta_1 + x_{i.}^{(1)}(\gamma_1 - \beta_1) + x_i^{(2)}\gamma_2 + v_i + u_{it} \end{aligned}$$

that shows that, to control for correlated fixed effects $E(x_i^{(1)}\alpha_i) \neq 0$, $E(x_i^{(2)}\alpha_i) \neq 0$, it suffices to add the control functions $x_{i.}^{(1)}$ to estimate consistently β_1.

Remark: Parameters γ have no economic meaning. There are sums of structural parameters β_1 and β_2 and control parameters, θ_1 and θ_2.
We can identify β_1, γ_1 and γ_2 where $\gamma_1 = \beta_1 + \theta_1$ and $\gamma_2 = \beta_2 + \theta_2$ hence β_1, θ_1 and $\beta_2 + \theta_2$ are identified.
β_2 is not identified except if we use instruments (see below).

3 Uncorrelated Random Effects

3.1 Identification and Test

Remark 2: We have an estimator of β_1 in the within dimension and an estimator of $\beta_1 + \theta_1$ in the between dimension, recombine both to obtain an estimator of θ_1 :

$$\hat{\theta}_1 = \hat{\gamma}_1^{(b)} - \hat{\beta}_1^{(w)}$$

The asymptotic variance-covariance matrix of this estimator is the sum of the asymptotic variance-covariance matrices of these two estimators since they are obtained in the between and within dimensions which are orthogonal. We can then test the hypothesis that it is equal to 0.

There is a direct procedure since the model can be rewritten as a function of $x_{it}^{(1)}$, $x_{i.}^{(1)}$, $x_i^{(2)}$ and using $\gamma_2 = \beta_2$ as:

$$y_{it} = x_{it}^{(1)}\beta_1 + x_{i.}^{(1)}(\gamma_1 - \beta_1) + x_i^{(2)}\beta_2 + v_i + u_{it}$$

By definition $\theta_1 = \gamma_1 - \beta_1$. The OLS estimators in this equation are $\hat{\beta}_1^{(w)}$, $\hat{\theta}_1$ and $\hat{\beta}_2^{(b)} = \hat{\gamma}_2^{(b)}$. The test of $H_0 : \theta_1 = 0$ is thus the usual Wald test.

Remark: We can find in the literature a popular procedure called Hausman test (Hausman et Taylor, 1981) but the statistic used in this procedure is *numerically* equal to the Wald statistic that we here present (Arellano, 1993). Moreover, this one extends easily to the heteroskedastic and unbalanced cases.

Interpretation of $\theta_1 = 0$:

$$\theta_1 = 0 \Leftrightarrow E(x_{i.}^{(1)\prime}(\alpha_i - x_i^{(2)}\theta_2)) = 0$$

There is no correlation between individual effects and explanatory variables.
The true model becomes:

$$y_{it} = x_{it}^{(1)}\beta_1 + x_i^{(2)}\gamma_2 + v_i + u_{it}$$

under hypotheses H_1, H_2 and H_3'.

Remark 1: For simplicity, no correlation between individual effects and variables constant over time, $x_i^{(2)}$. Set $\theta_2 = 0$ and $E(x_i^{(2)\prime}v_i) = 0$ and $\gamma_2 = \beta_2$.
Remark 2: The disturbance $v_i + u_{it}$ is not correlated, by construction, to explanatory variables and the OLS estimator is now consistent.
Remark 3: The between and within estimator are consistent as well.
Remark 4: Some efficiency gains to be expected since the within and between dimensions are not independent and thus GLS estimation is recommendable (Balestra and Nerlove, 1966). (*Random effect model).*

3.2 Model and Estimation

Hypotheses H_1, H_2 to H_3' and $\theta_1 = 0$.

The error-component disturbance has a one factor structure:

$$\varepsilon_{it} = v_i + u_{it}.$$

Hypotheses H_2 and H_3' imply:

$$E(\varepsilon_i \varepsilon_i' \mid x_i) = \Sigma = \begin{pmatrix} \sigma_v^2 + \sigma_u^2 & \sigma_v^2 & \dots & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 + \sigma_u^2 & \sigma_v^2 & \vdots \\ \vdots & \sigma_v^2 & \ddots & \sigma_v^2 \\ \sigma_v^2 & \dots & \sigma_v^2 & \sigma_v^2 + \sigma_u^2 \end{pmatrix}$$

The second order moment matrix of disturbances is equal to:

$$E(\varepsilon\varepsilon' \mid X) = \sigma_u^2 (W + \frac{1}{\rho^2} B)$$

where

$$\rho = \frac{1}{\sqrt{T}} \left(\frac{\sigma_u^2}{\frac{\sigma_u^2}{T} + \sigma_v^2} \right)^{1/2} \text{ varies between 0 and 1.}$$

(See Proof)

Remark: This coefficient tends to 0 when T tends to infinity.

Definition: The random effect model under hypotheses H_1, H_2 and H_3' is the general linear model:

$$Y = X_1\beta_1 + BX_2\gamma_2 + \varepsilon$$

$$E(X'\varepsilon) = 0$$

$$E(\varepsilon\varepsilon' \mid X) = \sigma_u^2(W + \frac{1}{\rho^2}B) = \sigma_u^2\Omega$$

3.2.1 Generalized Least Squares

Assume that ρ is known and that $H = W + \rho B$. We have:

$$H.\Omega.H' = I$$

so that the sphericized model is:

$$(W + \rho B)Y = (W + \rho B)X_1\beta_1 + \rho BX_2\beta_2 + H\varepsilon$$

where $E((H\varepsilon)(H\varepsilon)') = \sigma_u^2 I$.
If we are only interested by β_1, we can use the Frish-Waugh theorem by premultipying this equation by the projector on the space orthogonal to variables BX_2, that we denote M_{BX_2}. Note also that

$M_{BX_2}W = W$ since BX_2 varies in the between dimension only. Thus:

$$(W + \rho M_{BX_2}B)Y = (W + \rho M_{BX_2}B)X_1\beta_1 + M_{BX_2}H\varepsilon$$

The GLS estimator is thus:

$$\hat{\beta}_1^{(GLS)} = (X_1'(W + \rho^2 BM_{BX_2}B)X_1)^{-1}X_1'(W + \rho^2 BM_{BX_2}B)Y$$

Property: The GLS estimator is a weighted mean of within and between estimators and thus combine the information in the within and between dimensions.

$$\hat{\beta}_1^{(GLS)} = \Lambda.\hat{\beta}_1^{(w)} + (I - \Lambda)\hat{\beta}_1^{(b)}$$

(See Proof).

Remark: ρ tends to 0 when T tends to infinity. Numerically, $\hat{\beta}_1^{(GLS)}$ tends to the within estimator when T tends to infinity.
As we generally do not know ρ, we adopt feasible GLS.

3.2.2 Feasible GLS

Let a consistent estimator of ρ, $\hat{\rho}$ and replace the unknown parameter by its estimator:

$$\hat{\beta}_1^{(FGLS)} = (X_1'(W + \hat{\rho}^2 BM_{BX_2}B)X_1)^{-1}X_1'(W + \hat{\rho}^2 BM_{BX_2}B)Y$$

Consistent estimator of ρ : Use the two consistent estimators within and between.
Within dimension: the variance of disturbances is :

$$E(W\varepsilon\varepsilon' W \mid X) = \sigma_u^2 W$$

which enables to construct a consistent estimator of σ_u^2 using the residuals of within estimation:

$$\hat{\sigma}_u^2 = \frac{\hat{\varepsilon}^{(w)\prime} W \hat{\varepsilon}^{(w)}}{N(T-1)}.$$

Between dimension:

$$E(B\varepsilon\varepsilon' B \mid X) = \frac{\sigma_u^2}{T\rho^2} B$$

which permits to construct a consistent estimator of $\frac{\sigma_u^2}{T\rho^2}$ using the residuals of the between estimation:

$$T\widehat{\sigma_v^2 + \sigma_u^2} = \frac{\widehat{\sigma_u^2}}{\rho^2} = \frac{\hat{\varepsilon}^{(b)\prime} B \hat{\varepsilon}^{(b)}}{N}.$$

The consistent estimator of ρ^2 is thus given by:

$$\hat{\rho}^2 = \frac{\hat{\sigma}_u^2}{\widehat{\frac{\sigma_u^2}{\rho^2}}}$$

Remark: This estimator does not necessarily dominate other estimators (as the within and between estimators) in finite samples. Its properties are asymptotic only.

3.3 Asymptotic properties of estimators

Property 1: GLS, FGLS and the within estimators are consistent when $n \to \infty$, or $T \to \infty$, or both. They are asymptotically equivalent when $T \to \infty$.
(See Proof)
Property 2: OLS, and the between estimators are consistent when $n \to \infty$, or when $n \to \infty$ and $T \to \infty$, but not when $T \to \infty$.
(See Proof)

4 Instrumental variables, heteroskedasticity and serial correlation

Random effects & Fixed effects: Correlation betweeen α_i and $x_{i.}^{(1)}$ and $x_i^{(2)}$ or not.
Intermediate cases? Other orthogonality conditions?

4.1 Instrumental Variables

4.1.1 Hausman & Taylor (1981)

Suppose that explanatory variables, varying or not over time, are of two types, the ones which are correlated with individual effects and the ones which are not:

$$x_{it}^{(1)} = (s_{it}^{(1)}, z_{it}^{(1)}) \quad x_i^{(2)} = (s_i^{(2)}, z_i^{(2)})$$

under hypothesis H_5:

$$E(s_{it}^{(1)\prime}\alpha_i) = h^{(1)} \quad E(s_i^{(2)\prime}\alpha_i) = h^{(2)}$$

$$E(z_{it}^{(1)\prime}\alpha_i) = 0 \quad E(z_i^{(2)\prime}\alpha_i) = 0$$

The Hausman et Taylor method consists in estimating the equation:

$$y_{it} = x_{it}^{(1)}\beta_1 + x_i^{(2)}\beta_2 + \alpha_i + u_{it}$$

using a instrumental variable method where instruments are the variables uncorrelated with disturbance $\alpha_i + u_{it}$ under H_1 and H_5:

- variables $z_{it}^{(1)}$ and $z_i^{(2)}$
- individual means $z_{it}^{(1)}$, $z_{i.}^{(1)}$ and differences of explanatory variables $s_{it}^{(1)}$ with their individual means,

$s_{it}^{(1)} - s_{i.}^{(1)}$ since the correlation between $s_{it}^{(1)}$ and α_i is constant over time.

Order condition for identification: the number of instruments should be larger that the number of parameters. Assume that the number of variables in $z_{i.}^{(1)}$ is larger that the number of variables in $s_i^{(2)}$.

Estimation: Under homoskedasticity, estimate by 2SLS. As the model is over-identified, the efficient estimator is 3SLS where the correlation over time between disturbances, $\alpha_i + u_{it}$, as given by H_2 and H_3, is taken into account. It amounts to consider that the model is written for an individual as a system of T simultaneous equations.

4.1.2 Gaining precision

Amemiya and MaCurdy (1986): other valid instruments are explanatory variables $z_{it'}^{(1)}$ at all dates t' under hypotheses H_1 and H_5.

Breusch, Mizon and Schmidt (1989): differences to individual means of explanatory variables $s_{it'}^{(2)} - s_i^{(2)}$ at all dates t' are also valid instruments.

Arellano et Bover (1995) decomposes the model into its within and between dimension:

$$\begin{aligned} y_{it} - y_{i.} &= (x_{it} - x_{i.})\beta + u_{it} - u_{i.} \\ y_{i.} &= x_{i.}\beta + \alpha_i + u_{i.} \end{aligned}$$

The instruments in the first equation are:

for any t', $x_{it'}^{(1)}$

In the second equation:

$$z_{i.}^{(1)}, z_i^{(2)}$$

4.2 Heteroskedasticity and Serial Correlation

Assess its impact on known estimators or change estimation method (FGLS and GMM). Consider:

$$y_{it} = x_{it}\beta_i + \alpha_i + u_{it}$$

where slopes are random:

$$\beta_i = \beta + \zeta_i \text{ where } E(\zeta_i \mid x_i) = 0.$$

It implies that:

$$\begin{aligned} y_{it} &= x_{it}\beta + \alpha_i + x_{it}\zeta_i + u_{it} \\ &= x_{it}\beta + \alpha_i + \tilde{u}_{it}. \end{aligned}$$

Hypothesis H1 becomes:

$$\begin{aligned} E(x_{it}'\tilde{u}_{it}) &= E(x_{it}'(x_{it}\zeta_i + u_{it})) \\ &= E(x_{it}'x_{it}E(\zeta_i \mid x_i)) = 0. \end{aligned}$$

However:

$$E(\tilde{u}_{it}^2 \mid x_i) = E(u_{it}^2 \mid x_i)$$
$$+E(x_{it}\zeta_i\zeta_i' x_{it}' \mid x_i) + 2E(u_{it}\zeta_i' x_{it}' \mid x_i) = 2\sigma^2 + \Phi(x_i)$$

Similarly for $E(\tilde{u}_{it}\tilde{u}_{it}' \mid x_i)$.
Thus there is heteroskedasticity and serial correlation of unknown form.

Does not affect consistency of the various estimators but affect the efficiency property of feasible GLS and the standard errors of all estimators.
Correct standard errors are the ones robust to heteroskedasticity by clusters (see Chapter on fixed effects)

4.3 Estimating covariance structures

4.3.1 Restrictions

Other types of heteroskedasticity can stem from structural reasons. For instance that u_{it} is postulated to have a MA(1) structure:

$$u_{it} = \theta\eta_{it-1} + \eta_{it}.$$

where η_{it} is white noise.

The variance-covariance matrix Σ is a function of θ and other parameters (See Proof):

$$\Sigma = \Sigma(\sigma_v^2, \sigma_\eta^2, \theta).$$

Choose any consistent estimator of β and construct a estimator $\hat{\Sigma}$ of the variance-covariance matrix of errors:

$$\hat{\varepsilon}_{it} = y_{it} - x_{it}\hat{\beta}$$

that is of elements:

$$\hat{\Sigma}_{tt'} = \frac{1}{n}\sum \hat{\varepsilon}_{it}\hat{\varepsilon}_{it'}.$$

Notations: There are $T(T+1)/2$ independent elements in Σ. By vectorialization of the upper triangular part of Σ, $\Lambda = vech(\Sigma)$, we can write the restriction as:

$$\Lambda = \Lambda(\sigma_v^2, \sigma_\eta^2, \theta).$$

4.3.2 Minimum distance estimation

Two step estimation: Estimate first $\hat{\Lambda}_n = vech(\hat{\Sigma})$ as derived above. Suppose that:

$$\sqrt{n}(\hat{\Lambda}_n - \Lambda) \underset{n\to\infty}{\overset{d}{\rightsquigarrow}} N(0, V)$$

where an element of V is $v_{tt'} = V(\varepsilon_{it}\varepsilon_{it'}) = E(\varepsilon_{it}^2\varepsilon_{it'}^2) - (E\varepsilon_{it}\varepsilon_{it'})^2$ and would then be consistently estimated (under some conditions on the higher moments) by:

$$\hat{v}_{tt'} = \frac{1}{n}\sum(\hat{\varepsilon}_{it}^2 - \bar{\varepsilon}_t^2)(\hat{\varepsilon}_{it'}^2 - \bar{\varepsilon}_{t'}^2).$$

Rewrite the estimation problem as:

$$\hat{\Lambda}_n = \Lambda + (\hat{\Lambda}_n - \Lambda) = \Lambda(\sigma_v^2, \sigma_\eta^2, \theta) + \frac{\hat{R}_n}{\sqrt{n}}$$

denoting $\hat{R}_n = \hat{\Lambda}_n - \Lambda$ where $\hat{R}_n$ is approximately normally distributed.

The optimal minimum distance estimator is given by:

$$\min_{(\sigma_v^2,\sigma_\eta^2,\theta)} (\hat{\Lambda}_n - \Lambda(\sigma_v^2, \sigma_\eta^2, \theta))'(\hat{R}_n)^{-1}(\hat{\Lambda}_n - \Lambda(\sigma_v^2, \sigma_\eta^2, \theta))$$

Remark: The small sample properties might not be good (Altonji and Segal, 1996). We can do various things. Replace it by a continuously updated estimation (Hansen, Heaton, Yaron, 1996) i.e. at each step of the algorithm k use a weight matrix taking also into account $(\hat{\Lambda}_n - \Lambda_k)$, generalized empirical likelihood methods (Newey and Smith, 2004), or maximum likelihood methods (Alvarez and Arellano, 2004).

4.4 Measurement errors

True model:

$$y_{it} = \alpha_i + x_{it}^{*}\beta + u_{it},\ i = 1, ., n; t = 1, ., T$$

where all variables are assumed to vary over time.
Measurement: Observed variables are:

$$x_{it} = x_{it}^{*} + \tau_{it},$$

where τ_{it} is a classsical measurement error. It is uncorrelated with $u_{it'}, x_{it'}^{*}$ across any period t'.

Within estimator: It is biased since:

$$plim_{n\to\infty}\hat{\beta}_w = (I - (plim_{n\to\infty}\frac{X'WX}{n})^{-1}plim_{n\to\infty}\frac{\tau'W\tau}{n})\beta$$

(See Proof).
The between estimator can be expressed in the same way.

Simplification: Single dimensional x_{it}.

$$plim_{n\to\infty}\frac{X'WX}{n} = E(\sum_{t=1}^{T}(x_{it} - x_{i.})^2) \quad plim_{n\to\infty}\frac{\tau'W\tau}{n} = E(\sum_{t=1}^{T}(\tau_{it} - \tau_{i.})^2) = (T-1)\sigma_\tau^2$$

Between estimator:

$$plim_{n\to\infty}\frac{X'BX}{n} = E(Tx_{i.}{}^2) \quad plim_{n\to\infty}\frac{\tau'B\tau}{n} = E(T\tau_{i.}{}^2) = \sigma_\tau^2$$

Suppose that $x_{it}^* = z_i + w_{it}$ and w_{it} is stationnary (constant variance):

$$E(\sum_{t=1}^{T}(x_{it} - x_{i.})^2) = (T-1)(\sigma_w^2 + \sigma_\tau^2) \quad E(Tx_{i.}{}^2) = (TE(z_i^2) + \sigma_w^2 + \sigma_\tau^2)$$

Conclusion: The attenuation bias is much lower in the case of the between estimator **but** the bias due to the random effect assumption is much larger.

Solution: As τ_{it} and $\tau_{it'}$ are independent, $x_{it'}$ is a good instrument in the within regression.

4.5 Extensions

Sometimes, distribution functions of α_i and u_{it} are needed: transition probabilities between income classes for instance in an analysis of income inequality.
Idea: Use deconvolution techniques to estimate the distribution of α_i and u_{it} using the distribution of residuals:

$$\hat{\varepsilon}_{it} = y_{it} - x_{it}\hat{\beta}_n$$

where $\hat{\beta}_n$ is any consistent estimate.

See Horowitz and Markatou (1996) for deconvolution techniques using characteristic functions.

5 Summary and conclusion

In the static model:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

Two conclusions:

- Pay attention to the correlation between α_i and x_{it}. For instance write:

  $$y_{it} = x_{it}\beta + x_{i.}\theta + v_i + u_{it}$$

  test that $\theta = 0$ and proceed ...

- We never questionned hypothesis $H1$, $E(x'_{it}u_{i\tau}) = 0$. In dynamic models it is questionable since $x_{i\tau}$ could be the result of shocks affecting y_{it} next period.