1 Introduction and motivation

- Panel data (or longditudinal data) refers to a cross-section repeatedly sampled over time, but where the same economic agent has been followed throughout the period of the sample.
- Examples.
  - firm or company data
  - longditudinal data on patterns of individual behaviour over the life-cycle.
  - comparative country-specific macroeconomic data over time.
- Common feature:
  - the sample of individuals N is typically relative large
  - the number of time periods T is generally short.

- Why use panel estimation methods? Can answer questions not possible either from a cross-section context or with a pure time series.
- Greene (1991) we observe 50 per cent of a cohort of women to work. Two possible interpretations
  - 50 per cent of women work on average each period, or .
  - *the same* 50 per cent of women may work each period. Different interpretations, different implications for policy.
- There are nevertheless difficulties inherent in data sources with a longditudinal element.

  (a) *attrition*

  (b) *non-randomness of the sample*.

2 Why use panel data methods?

- increased precision of regression estimates
- the ability to control for individual fixed effects
- the ability to model temporal effects without aggregation bias

3 Fixed effects panel data models

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \qquad (1)$$

for $i = 1, ..., N$ individuals over $t = 1, ..., T$ time periods.

- Model includes
  - an individual effect α_i (constant over time).
  - marginal effects β for x_{it} (common across i and t).

3.1 The pooled Ordinary Least Squares (OLS) estimator

- the simplest approach to the estimation.
- individual effects α_i are fixed and common across economic agents, such that $\alpha_i = \alpha$ for all $i = 1, ..., N$. .
- OLS produces consistent and efficient estimates of α and β.

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$$

$$\widehat{\beta} = \frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{y}_{it}}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}^{2}} \tag{2}$$

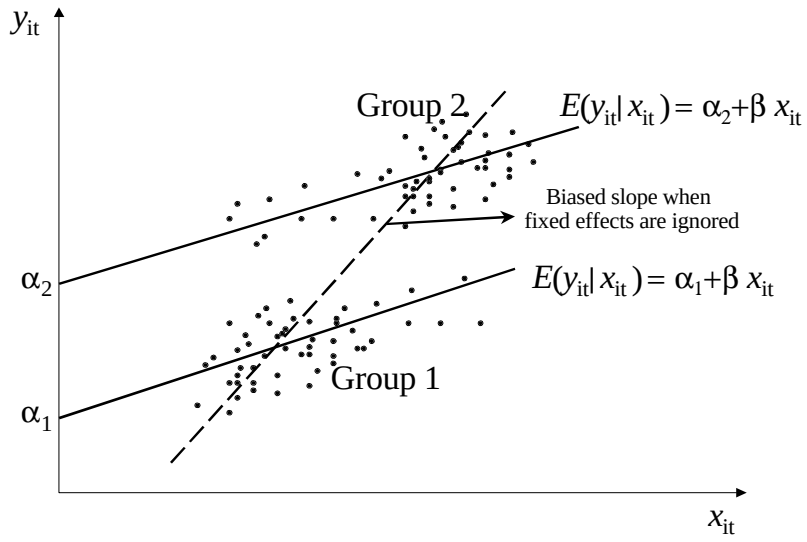
where

- $\overline{x} = (1/NT).\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}$ and
- $x_{it} = x_{it} - \overline{x}$ (similarly for y).

Notice

$$var(\widehat{\beta}) = \frac{var(u_{it})}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}^2} \qquad (3)$$

1.Bias from ignoring fixed effects



3.2 The Within-Groups (WG) estimator

- can be used if individual effects α_i are fixed but not common across $i = 1, ..., N$
- eliminates the fixed effect α_i by differencing

.

- Let $\overline{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ and
- $\overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$.
- Define

$$\begin{aligned} x_{it}^* &= x_{it} - \overline{x}_i \\ \text{and } y_{it}^* &= y_{it} - \overline{y}_i, \end{aligned}$$

- Then

$$\overline{y}_i = \alpha_i + \overline{x}_i' \beta + \overline{u}_i.$$

- Subtracting from (1) gives

$$y_{it} - \overline{y}_i = (\alpha_i - \alpha_i) + (x_{it} - \overline{x}_i)' \beta + (u_{it} - \overline{u}_i)$$

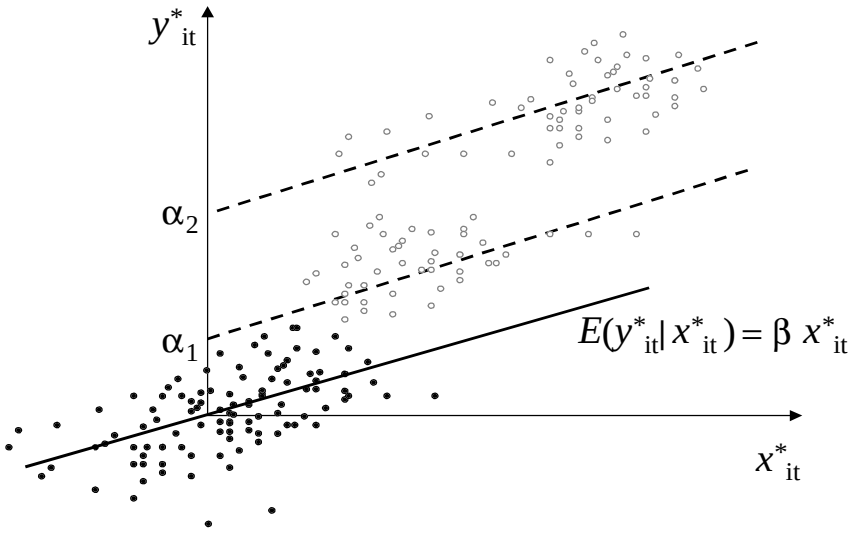
or

$$y_{it}^* = x_{it}^{*\prime} \beta + u_{it}^*. \tag{4}$$

- Hence,

$$\widehat{\beta}^{WG} = \frac{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}^* \widetilde{y}_{it}^*}{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}^{*2}} \tag{5}$$

2.The Within-Groups estimator



3.3 Variance of WG estimator

- $$\begin{aligned} S_{xx} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x})^2, \\ S_{xx}^{w} &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)^2 \\ S_{xx}^{b} &= \sum_{i=1}^{N}T(\overline{x}_i-\overline{x})^2 \end{aligned}$$

- Can show that

$$S_{xx} = S_{xx}^{w} + S_{xx}^{b}.$$

Given that $var(u_{it}^{*}) = (\frac{T-1}{T})var(u_{it})$, we have

$$var(\widehat{\beta}^{WG}) = \frac{var(u_{it}^{*})}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}^{2}} \tag{6}$$

$$= \frac{var(u_{it}^{*})}{S_{xx}-S_{xx}^{b}} \tag{7}$$

$$= \frac{(\frac{T-1}{T})var(u_{it})}{S_{xx} - S^b_{xx}} \qquad (8)$$

3.4 Drawback with the Within-Groups estimator

- eliminates time-invariant characteristics from a model of the form

$$y_{it} = \alpha_i + x'_{it}\beta + z'_i\delta + u_{it}$$

3.5 The Least Squares Dummy Variable (LSDV) estimator

- Define a series of group-specific dummy variables $d_{git} = \mathbf{1}(g = i)$.
- This gives

$$\begin{aligned} y_{it} &= \alpha_i + x'_{it}\beta + u_{it}, \\ &= \alpha_1 d_{1it} + \alpha_2 d_{2it} + ... + \alpha_N d_{Nit} + x'_{it}\beta + u_{it} \end{aligned} \quad (9)$$

- Estimate by standard OLS yielding $\widehat{\beta}_{LSDCV}$.
- A test for individual effects? Under the null,

$$\alpha_1 = \alpha_2 = ... = \alpha_N$$

- Test using subset-F statistic

$$F = \frac{R^2_{DV} - R^2_p}{1 - R^2_{DV}} . \frac{NT - N - k}{N - 1}$$

- Distributed $F_{N-1,NT-N-k}$ under the null of equality of α_i. .

3.6 The Two-Way Fixed Effects Model

.

$$y_{it} = \alpha_i + \gamma_t + x'_{it}\beta + u_{it},$$

where γ_t represents the (fixed) time effects

- Include time dummies $z_{sit} = \mathbf{1}(s = t)$ to give

$$\begin{aligned} y_{it} = \; & \alpha_1 d_{1it} + \alpha_2 d_{2it} + ... + \alpha_N d_{Nit} \\ & + g_2 z_{2it} + ... + g_T z_{Tit} + x'_{it}\beta + u_{it}. \end{aligned}$$

4 The random effects model (REM)

The fixed effects model is appropriate when differences between individual agents may reasonably be viewed simply as parametric shifts in the regression function itself. This might be considered reasonable if the cross-sectional used in estimation represents a broadly exhaustive sample of the population of economic agents, as might be the case in a study which covers a

full sample of countries, or in a study of the performance of firms in a particular industry, where the sample of firms represents a broadly complete coverage of those within the industry. If, on the other hand, the cross-section is drawn from a larger population (so that the sample of cross-sectional agents may not reasonably be considered exhaustive) then it may be more appropriate to view the individual-specific terms in the sample as randomly distributed effects across the full cross-section of agents. Defining $\alpha_i = \alpha + \tau_i$, where τ_i has a zero (unconditional) mean, this would suggest a random effects specification of the following form;

$$y_{it} = \alpha + x'_{it}\beta + u_{it} + \tau_i. \qquad (10)$$

Here, τ_i represents an individual disturbance which is fixed over time. The following assumptions relate to the random components in the model;

$$\begin{aligned}
E(u_{it}|\tau_i) &= 0 \\
E(u_{it}^2|\tau_i) &= \sigma_u^2 \\
E(\tau_i|x_{it}) &= 0 \text{ for all } i, t \\
E(\tau_i^2|x_{it}) &= \sigma_\tau^2 \\
E(u_{it}\tau_j) &= 0 \text{ for all } i, t, j \\
E(u_{it}u_{js}) &= 0 \text{ for } i \neq j \text{ or } t \neq s \\
E(\tau_i\tau_j) &= 0 \text{ for } i \neq j.
\end{aligned}$$

Among these assumptions, perhaps one of the more restrictive relates to the conditional expectation $E(\tau_i|x_{it})$, which is assumed to be zero for the simple random effects model. This may not be supportible, particularly in light of the fact that (10) does not contain any time-invariant characteristics specific to each individual in the sample (Examples: gender, education, parent's education), and ought at the very least to be tested.

4.1 The Generalised Least Squares (GLS) estimator

To estimate the linear random effects model (sometimes called the variance components or random components model) requires a Generalised Least Squares approach to deal with the more complex error structure inherent in (10) compared with the fixed effects model. To see this, consider the characteristics of the combined error term $w_{it} = u_{it} + \tau_i$. It is certainly true that $E(w_{it}) = 0$. However,

$$\begin{aligned} E(w_{it}^2) &= \sigma_u^2 + \sigma_\tau^2 \text{ for all } i, t \\ E(w_{it}w_{is}) &= \sigma_\tau^2 \text{ for all } t \neq s \\ E(w_{it}w_{js}) &= 0 \text{ for } i \neq j \text{ or } t \neq s \ . \end{aligned}$$

So, if we collect the T disturbances for individ-

ual i in a vector of the form $w_i = (w_{i1}, w_{i2}, ..., w_{iT})'$, we have that

$$E(w_i w_i') = \Omega,$$

where

$$\Omega = \begin{pmatrix} \sigma_u^2 + \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 & . & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_u^2 + \sigma_\tau^2 & \sigma_\tau^2 & . & \sigma_\tau^2 \\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_u^2 + \sigma_\tau^2 & . & \sigma_\tau^2 \\ & & . & . & . \\ \sigma_\tau^2 & \sigma_\tau^2 & \sigma_\tau^2 & . & \sigma_u^2 + \sigma_\tau^2 \end{pmatrix}.$$

For the full panel of observations, the covariance matrix of the NT vector of disturbances $w = (w_1, w_2, ..., w_N)'$ may be derived as

$$\begin{aligned} \underset{(NTxNT)}{V} &= \begin{pmatrix} \Omega & 0 & 0 & . & 0 \\ 0 & \Omega & 0 & . & 0 \\ 0 & 0 & \Omega & . & 0 \\ & & . & . & . \\ 0 & 0 & 0 & . & \Omega \end{pmatrix} \\ &= I_N \otimes \Omega \end{aligned} \tag{11}$$

where I_N is the identity matrix of dimension N and $\otimes$ represents the Kronecker product. Let Y represent a stacked vector of y_{it} formed in a similar fashion to w (with a similar structure for X). The full system

$$Y = X\beta + w \tag{12}$$

may therefore be estimated by Generalised Least

Squares, given the structure of the covariance matrix $\dot{V}$. Generally, GLS estimation of a regression of the form (12) requires a transformation to remove the non-standard structure of the covariance matrix $E(ww') = V$. We define the weight matrix $P = V^{-\frac{1}{2}}$, and transform (12) by premultiplication, to give

$$PY = PX\beta + Pw$$

or

$$Y^* = X^*\beta + w^*.$$

Note now that

$$\begin{aligned} E(w^*w^{*\prime}) &= E(Pww'P) \\ &= P.E(ww')P \\ &= P.V.P \\ &= I_{NT} \end{aligned}$$

which has common variances across i and T. So, with knowledge of P the GLS estimators of the regression function (12) may be derived as

$$\widehat{\beta}_{GLS} = (X'V^{-1}X)^{-1}.X'V^{-1}Y. \qquad (13)$$

Do recall, however, that we were required to assume that $E(\tau_i|x_{it}) = 0$ in the If $E(\tau_i|x_{it}) \neq 0$ then the GLS estimator is not consistent.

For the random effects model, one can generate a specific form for the weight matrix $P = V^{-\frac{1}{2}}$. Given that $V^{-\frac{1}{2}} = I_N \otimes \Omega^{-\frac{1}{2}}$, we can

rewrite (11) as

$$\begin{aligned} V &= I_N \otimes \Omega \\ &= \sigma_u^2 I_N + \sigma_\tau^2 \mathbf{i}\mathbf{i}' \end{aligned}$$

where $\mathbf{i}$ represents an N-vector of ones. This allows us to derive the form of $\Omega^{-\frac{1}{2}}$ as

$$\Omega^{-\frac{1}{2}} = I_N - \frac{\theta}{T}.\mathbf{i}\mathbf{i}'$$

where

$$\theta = 1 - \frac{\sigma_u}{T(\sigma_\tau^2 + \sigma_u^2)^{\frac{1}{2}}}.$$

So, the appropriate transformation for the random effects model is to premultiply each $y_i = (y_{i1}, .., y_{iT})'$ by $\Omega^{-\frac{1}{2}}$ to give

$$\begin{aligned} y_i^* &= \Omega^{-\frac{1}{2}}.y_i \\ &= \begin{bmatrix} y_{i1} - \theta\overline{y}_i \\ y_{i2} - \theta\overline{y}_i \\ . \\ . \\ y_{iN} - \theta\overline{y}_i \end{bmatrix}, \end{aligned}$$

with similar transformations to generate each x_i^*.

4.2 The Breusch-Pagan Lagrange Multiplier test

The Breusch-Pagan LM statistic provides a test of the random effects model against the pooled OLS model given by (2). The specific hypothesis under investigation is the following:

$$H_0 \ : \ \sigma_\tau = 0$$
$$H_A \ : \ \sigma_\tau \neq 0.$$

From inspection of (11) one can see that $V = \sigma_\tau^2 . I_{NT}$ under the null $\sigma_\tau = 0$, so that the REM reduces to a pooled OLS regression. The test of this hypothesis, based on OLS residuals $\widehat{u}_{it}$ from the pooled regression model, requires the LM statistic

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \widehat{u}_{it} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{u}_{it}^2} - 1 \right]^2 .$$

Under the null, this statistic should be distributed as a χ_1^2.

4.3 The Hausman Test

We may be interested in comparing directly the random effects estimator $\widehat{\beta}_{GLS}$ with the fixed

effects estimator $\widehat{\beta}_{LSDCV}$. As noted above, in the presence of correlation between the regressors x_{it} and individual effects τ_i the GLS estimator is inconsistent, whilst the OLS estimates $\widehat{\beta}_{LSDCV}$ are consistent. If $E(\tau_i|x_{it}) = 0$ on the other hand, the GLS estimator is consistent and efficient whilst the OLS estimator is consistent but inefficient. This motivated a test proposed by Hausman (1978), who constructed a test based on the difference between $\widehat{\beta}_{GLS}$ and $\widehat{\beta}_{LSDCV}$. He noted that, under the null, the variance of the difference $\widehat{\beta}_{GLS} - \widehat{\beta}_{LSDCV}$ may be derived as

$$\begin{aligned} var(\widehat{\beta}_{GLS} - \widehat{\beta}_{LSDCV}) &= var(\widehat{\beta}_{GLS}) + var(\widehat{\beta}_{LSDCV}) - cov(\\ &= var(\widehat{\beta}_{GLS}) - var(\widehat{\beta}_{LSDCV}) \\ &= \Sigma, \text{ say,} \end{aligned}$$

since

$$cov(\widehat{\beta}_{GLS} - \widehat{\beta}_{LSDCV}, \widehat{\beta}_{LSDCV}) = cov(\widehat{\beta}_{GLS}, \widehat{\beta}_{LSDCV}) - var(\widehat{\beta}$$

The Hausman test of the null of no correlation can therefore be conducted using the Wald statistic

$$W = (\widehat{\beta}_{GLS} - \widehat{\beta}_{LSDCV})' \widehat{\Sigma}^{-1} (\widehat{\beta}_{GLS} - \widehat{\beta}_{LSDCV})$$

which is distributed as a chi-squared with k degrees of freedom under the null, k being the number of regressors in x_{it}.