TSE 23rd January 2019

Applied Econometrics for Development: Panel Estimation I (Theory)

Ana GAZMURI

Paul SEABRIGHT

Outline of Presentation

- Unobserved heterogeneity and its consequences
- Fixed and random effects
- Strict exogeneity
- Random effects estimation
- Fixed effects estimation
- Choosing between fixed and random effects estimation

Unobserved heterogeneity

- Suppose the outcome and the explanatory variables are observed over a number of periods *t*= 1,.....T and that the true underlying causal process is given by:

$$(1) \qquad y_{it} = \beta_0 + \mathbf{x}_{it}\beta + c_i + u_{it}$$

where $\mathbf{x}_{it}$ is 1 X K, c_i is an *unobserved* random variable (unobserved to the researcher, that is) which is constant over time. Different individual values of c_i can be considered different independent draws of this random unobserved variable, while the u_{it} are i.i.d. with mean zero.

OLS estimation of β is consistent iff

$$(2) \qquad \mathrm{E}\left(\mathbf{x}_t' c\right) = 0 \qquad \forall t$$

What if equation (2) does not hold?

- Taking first differences (note the constant term drops out):

$$(3) \qquad \mathrm{y}_{it} - \mathrm{y}_{it-1} = \left(\mathbf{x}_{it} - \mathbf{x}_{it-1}\right)\beta + \left(u_{it} - u_{it-1}\right)$$

$$\text{Write } \Delta \mathrm{y}_{it} = \mathrm{y}_{it} - \mathrm{y}_{it-1},\ \Delta\mathbf{x}_{it} = \left(\mathbf{x}_{it} - \mathbf{x}_{it-1}\right),\ \Delta u_{it} = \left(u_{it} - u_{it-1}\right)$$

- Then OLS estimation of equation (3) yields consistent estimates of β if

$$(4) \qquad \mathrm{E}\left(\Delta\mathbf{x}_t'\Delta u_t\right) = 0 \qquad \forall \mathrm{t}$$

and

$$(5) \qquad \operatorname{rank} \mathrm{E}\left(\Delta\mathbf{x}_t'\Delta\mathbf{x}_t\right) = K \qquad \forall \mathrm{t}$$

Notice what is implied by equation (4)

- Writing it out in full

$$(4') \qquad \mathrm{E}\left(\mathbf{x}_t' u_t\right) + \mathrm{E}\left(\mathbf{x}_{t-1}' u_{t-1}\right)$$

$$+\mathrm{E}\left(\mathbf{x}_t' u_{t-1}\right) + \mathrm{E}\left(\mathbf{x}_{t-1}' u_t\right) = 0 \qquad \forall t$$

- Notice that this includes cross-temporal conditional expectations: it's not enough that the simultaneous conditional expectations are zero

- This will become important when we consider dynamic processes – it rules out processes in which the $\mathbf{x}_t$ variables include lagged values of *y*

Implications of equation (5)

- Equation (5) will not hold if some elements of x_t are constant over time for all members of the population

- Why? Then Δx_t will contain elements that are identically zero, so the matrix will no longer have full rank

Fixed and random effects

- The variable c_i is a random variable in the sense that it varies between individuals *i* in ways that are unknown to the researcher prior to the estimation

- The terms “fixed effects” and random effects” do not refer to different degrees of randomness in the c_i but rather to different distribution conditions on c_i which can be assumed in order to estimate the beta coefficients

- Random effects estimation requires a strong mean independence assumption:

$$(6) \qquad \mathrm{E}\left(c_i \middle| \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}\right) = \mathrm{E}\left(c_i\right)$$

Fixed and random effects (II)

- Random effects estimation therefore considers *all* the correlation between the *y* and **x** variables in the sample to be informative about their true relation

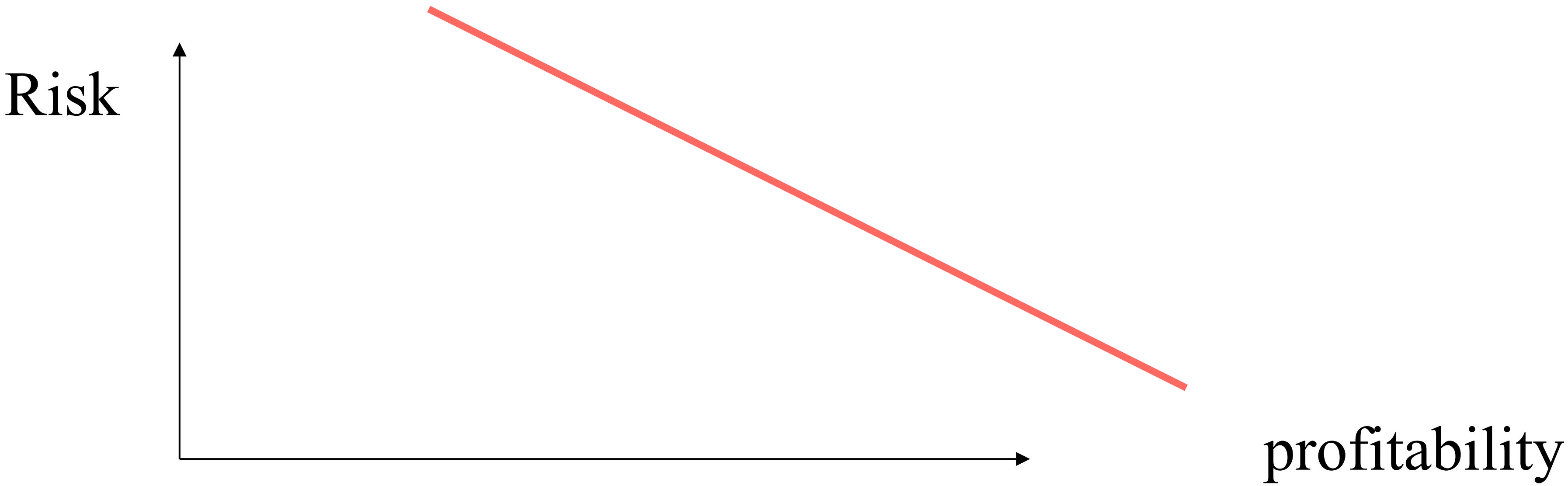
- Fixed effects estimation places no restrictions on the conditional distribution of *c* given the **x** variables, except the constancy of c_i over time for all *i*.

- The result is that only the correlation *over time* between the *y* and **x** variables is treated as informative about the true relation

- There are intermediate possibilities, notably correlated random effects, which allow for specific forms of dependence between the *c* and x variables

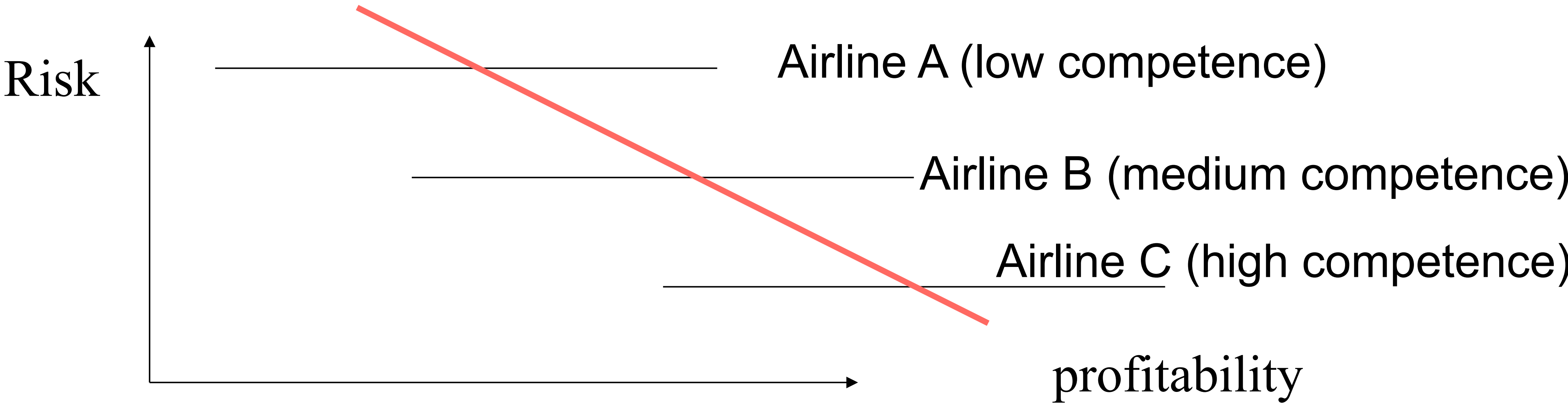
An example: the correlation between profitability and accident risk in airlines (Rose, JPE 1990)

OLS estimation yields an apparently negative relation – does this imply that reduced profitability leads airlines to take more risks?

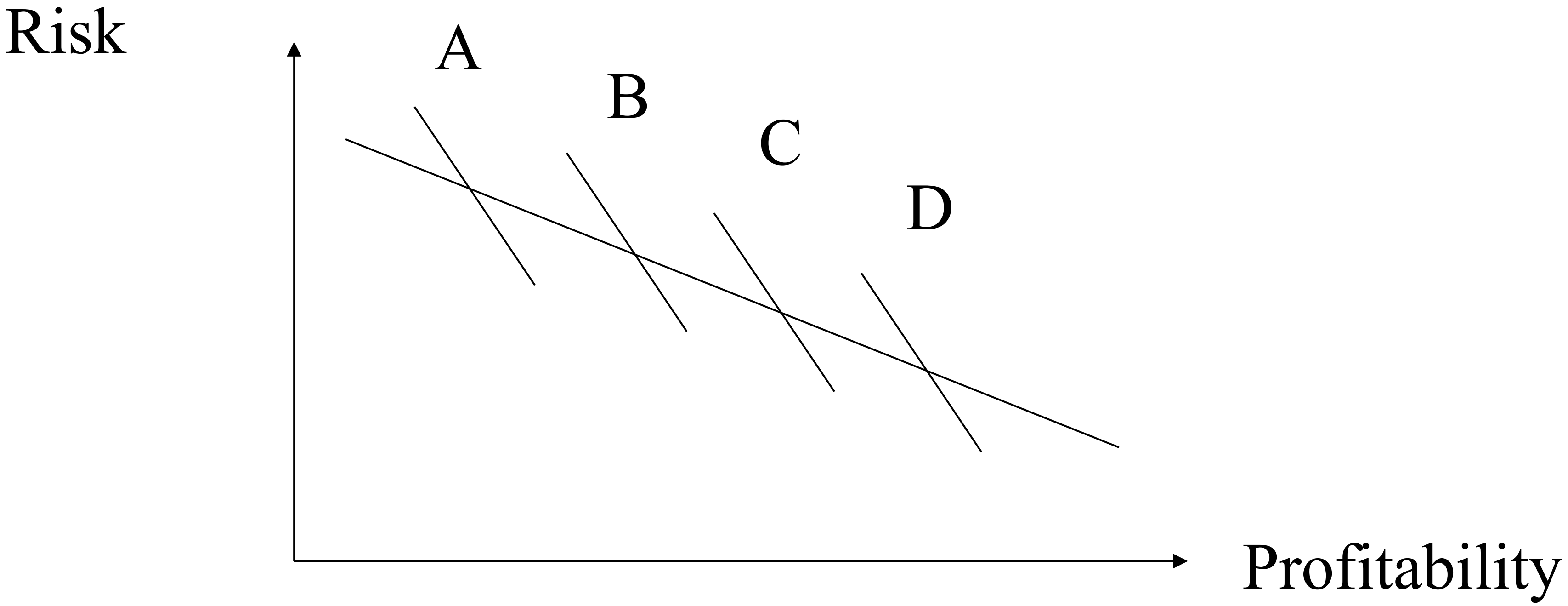


An example: the correlation between profitability and accident risk in airlines (Rose, JPE 1990)

Could the negative correlation between profitability and accidents be due to the (unobserved) competence of managers varying between airlines?



In fact the fixed effects go the other way.....



- Differences in accident rates not due to managerial competence but to cycles of profitability of an airline

Strict exogeneity

- The assumption of strict exogeneity of the **x** variables conditional on the unobserved effect can be written

$$(7) \qquad \mathrm{E}\left(y_{it} \middle| \mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i\right) = \mathrm{E}\left(y_{it} \middle| \mathbf{x}_{it},c_i\right) = \mathbf{x}_{it}\beta + c_i$$

- Which means that values of **x** in other time periods than *t* have no effect on the dependent variable in *t* once x_{it} and c_i are controlled for

- This will not hold if the determinants of *y* include its own lagged value (as in many wage equations, for instance), since then

$$(8) \qquad \mathrm{E}\left(u_{it} \middle| \mathbf{x}_{it+1}\right) \neq 0 \quad \text{since } \mathrm{y}_{it} \text{ is an element of } \mathbf{x}_{it+1}$$

Random effects estimation

- RE estimation requires both strict exogeneity of the x variables and the orthogonality of c_i and x_{it} for all t, ie that

$$(9) \qquad \mathrm{E}\left(c_i \middle| \mathbf{x}_i\right) = \mathrm{E}\left(c_i\right) = 0$$

- Note that if equation (9) holds without equation (7) we can still use pooled OLS estimation, but when equation (7) holds as well we can use a Generalized Least Squares (GLS) estimation which is more efficient. It exploits the known error structure and specifically the serial correlation of errors for each observation.

Random effects estimation (II)

- From equation (1), suppressing the constant, we have

$$(10) \qquad \text{y}_{it} = \mathbf{x}_{it}\beta + v_{it} \;\text{ where } \text{v}_{it} = c_i + u_{it}$$

- The model for all t periods can be written

$$(11) \qquad \mathbf{y}_i = \mathbf{x}_{it}\beta + \mathbf{v}_i \;\text{ where } \mathbf{v}_i = c_i\mathbf{j}_T + \mathbf{u}_i$$

- where $\mathbf{j}_T$ is the T X 1 vector of ones.
- If the u_i have a constant variance over time, and are serially uncorrelated, we can derive the form of the variance matrix

Random effects estimation (III)

- The variance matrix of v_i has the random effects structure:

$$(12) \qquad \mathbf{\Omega} = \mathrm{E}\left(\mathbf{v}_i \mathbf{v}_i'\right) = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \ldots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \ldots & \\ \ldots. & \ldots & \ldots & \sigma_c^2 \\ \sigma_c^2 & \ldots & \ldots & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$

Random effects estimation (IV)

- For efficient estimation the *conditional* variance matrix of v_i has to be constant (conditional on the x_i, that is):

$$(13a) \qquad \mathrm{E}\left(\mathbf{u}_i\mathbf{u}_i' \middle| \mathbf{x}_i, c_i\right) = \sigma_u^2 \mathbf{I}_T$$

$$(13b) \qquad \mathrm{E}\left(c_i^2 \middle| \mathbf{x}_i\right) = \sigma_c^2$$

- Then, given consistent estimators $\hat{\sigma}_u^2, \hat{\sigma}_c^2$, the Random Effects Estimator is given by

$$(14) \qquad \hat{\beta}_{RE} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{y}_i\right)$$

Random effects estimation (V)

- In the absence of condition (13) – for instance, if there is heteroskedasticity, or if the idiosyncratic errors are serially correlated - the RE estimator is still consistent, but it will not be efficient, and hypothesis testing must be done using robust standard errors
- There is a more general Feasible GLS estimator than the RE estimator with better large sample properties but poor finite sample properties for low *N* (see Wooldridge section 10.4.3)
- Testing for the presence of unobserved heterogeneity is a test of

$$\mathrm{H}_0 : \sigma_c^2 = 0$$

Random effects estimation (VI)

- Notice that the presence of random effects implies serial correlation in the error terms v_i - but this is not the same as serial correlation in the idiosyncratic error terms u_i
- The serial correlation implied by random effects is not autoregressive – the correlation between v_{it} and v_{is} does not tend to zero as *s-t* becomes large
- One implication is that general tests of serial correlation (eg Breusch-Pagan) do not distinguish between the presence of random effects and serial correlation in the idiosyncratic error terms

Fixed effects estimation

- FE estimation (just like RE estimation) requires equation (7) - strict exogeneity of the explanatory variables conditional on c_i
- However, equation (9) does not have to hold – the c_i may be arbitrarily correlated with the x_i
- FE estimation uses only time-varying information in the data. From (1), averaging over t=1,….,T and suppressing the constant:

$$(15) \qquad \bar{y}_i = \bar{\mathbf{x}}_i \beta + c_i + \bar{u}_i$$

where $\bar{y}_i = \sum_{t=1}^{T} y_{it} / T, \bar{\mathbf{x}}_i = \sum_{t=1}^{T} \mathbf{x}_{it} / T, \bar{u}_i = \sum_{t=1}^{T} u_{it} / T$

Fixed effects estimation (II)

- We can now define the divergence of each variable from its mean over time:

$$(16) \qquad \ddot{y}_i = \ddot{\mathbf{x}}_i \beta + \ddot{u}_i$$

$$\text{where } \ddot{y}_i \equiv y_{it} - \bar{y}_i, \ddot{\mathbf{x}}_i \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i, \ddot{u}_i \equiv u_{it} - \bar{u}_i$$

- Then under the assumption of strict exogeneity, pooled OLS estimation of (16) is consistent, and the FE estimator (also called the *within* estimator)is the pooled OLS estimator of β in (16)

- Note that for the matrix of time-demeaned explanatory variables to have full rank, it must exclude variables constant over time

Fixed effects estimation (III)

- The Fixed Effects (within) estimator can be written:

$$(17) \qquad \hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}'_{it} \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}'_{it} \ddot{y}_{it} \right)$$

- We can also define the *between* estimator from equation (15):

$$(18) \qquad \hat{\beta}_{B} = \left(\sum_{i=1}^{N} \overline{\mathbf{x}}'_{it} \overline{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \overline{\mathbf{x}}'_{it} \overline{y}_{it} \right)$$

- This is consistent under equation (9) but less efficient than the Random Effects estimator (which is a weighted average of the within and between estimators)

Fixed effects estimation (IV)

- For FE estimation to be efficient, the conditional variance matrix of the explanatory variables must be constant (equation 13a)

- However, the standard errors have to be calculated a little differently from OLS standard errors (this is done automatically in Stata and similar packages under a FE option). To see why note that the covariance of errors at times *t* and *s* are given by:

$$(19) \qquad \mathrm{E}\left(\ddot{u}_{it}\ddot{u}_{is}\right) = \mathrm{E}\left(\left(u_{it} - \bar{u}_i\right)\left(u_{is} - \bar{u}_i\right)\right)$$

$$= \mathrm{E}\left(u_{it}u_{is}\right) - \mathrm{E}\left(u_{it}\bar{u}_i\right) - \mathrm{E}\left(u_{is}\bar{u}_i\right) + \mathrm{E}\left(\bar{u}_i^2\right)$$

$$= 0 - \sigma_u^2 / T + \sigma_u^2 / T - \sigma_u^2 / T < 0$$

Fixed effects estimation (V)

- Serial correlation or heteroskedasticity in the errors can often be a problem in FE estimation

- As with RE, there are Fixed Effects Generalized Least Squares Estimators (and associated robust variance matrix estimators), the specific estimators that are appropriate depending on the assumptions that can be made about the errors (see Wooldridge sections 10.5.4 and 10.5.5)

- FE estimation is often subject to large attenuation bias due to measurement error (which can contribute a larger proportion of within variation than between variation); this means FE often does not avoid need for instrumental variables estimation

Fixed effects estimation (VI)

- A special case of FE is difference-in-difference estimation, where T=2 and where the key intervention is a group-level variable (so the differencing is designed to remove a group-level fixed effect)

- A pioneering example was John Snow's examination of the effect of the Lambeth Water Company moving its water sourcing upriver to a less sewage-contaminated area; he compared its fall in death rates from cholera to those of the Southwark and Vauxhall Water Company which did not move

- Often hard to know whether this is the right counterfactual

Choosing between fixed and random effects estimation

- FE estimation discards non-time varying information in **x**, which can lead to imprecise estimates if the main variation between observations *i* is cross-sectional rather than over time

- It also requires estimating an additional intercept for each extra observation, which can lead to problems of degrees of freedom if T is low. For fixed T and large N, fixed effects are not estimated consistently (this is called the “incidental parameters problem”).

- In linear panel estimation the β coefficients are still estimated consistently but in non-linear applications (eg logit) joint estimation of the intercepts and the β coefficients typically leads to inconsistent estimates of β

Choosing between fixed and random effects estimation (II)

- Hausman has a test comparing FE and RE estimators, with RE as the null. This test assumes
  - Strict exogeneity under both null and alternative hypotheses
  - Equation (9) holds under the null, so RE is efficient
- The test compares FE and RE estimators of β coefficients on variables that vary over both *i* and *t* – FE does not estimate coefficients for those that vary only over *i*, and for those that vary only over *t*, FE and RE estimates are identical
- Rewrite equation (10) with **z** as vector of time-constant variables

$$(20) \qquad y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it} = \mathbf{z}_{it}\gamma + \mathbf{w}_{it}\delta + c_i + u_{it}$$

Choosing between fixed and random effects estimation (III)

- Then Hausman test-statistic is

$$(21) \qquad H = \left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right)' \left[\hat{A}(\hat{\delta}_{FE}) - \hat{A}(\hat{\delta}_{RE})\right]^{-1} \left(\hat{\delta}_{FE} - \hat{\delta}_{RE}\right)$$

where $\hat{A}$ are consistent estimates of asymptotic variance

- This is distributed as χ^2_M under the null, where M is the number of time- and individual-varying variables (equation 20 excludes pure time effects)

Choosing between fixed and random effects estimation (III)

- Most importantly, FE and RE estimators have different interpretations
- Consider a study in which the *i* observations are countries; the variable y_i represents national income and x_i represents an endowment of mineral resources.
- Then the RE estimate of β represents the extent to which countries that have higher mineral resources tend to have higher income
- The FE estimate represents the extent to which a country that discovers new mineral resources will have a higher income

Choosing between fixed and random effects estimation (IV)

- Think whether it would be more appropriate to use OLS, RE or FE for the following studies
  - A study of the effect of winning a lottery on an individual's increase in income over a 5-year period after the win
  - A study of the effect of fertiliser use on output among poor farmers
  - A study of the effect of vaccination on child health

- What underlying principles are guiding your choices?

TSE 23rd January 2019

Applied Econometrics for Development: Panel Estimation I (Theory)

Ana GAZMURI

Paul SEABRIGHT