Pooled cross section over time

- Sample taken every year, but the cross section changes every year
- Often the distribution of variables changes over time but observations are still independent
- Can be treated as normal cross section but dummy variable for each year should be included
- Most of so called natural experiments are in fact pooled cross sections
- Simplest case
  - year 1 and 2 (no treatment, treatment)

– control group A (no treatment in year 2)
– treatment group B (treatment in year 2)

- dummy variable $dB = \begin{cases} 0 & \text{if in group } A \\ 1 & \text{if in group } B \end{cases}$

- dummy variable $d2 = \begin{cases} 0 & \text{if year } 1 \\ 1 & \text{if year } 2 \end{cases}$

- Simplest equation

$$y = \beta_0 + \delta_0 d2 + \beta_1 dB + \delta_1 d2 \cdot dB + u$$

- δ_0 captures the effect of year for both groups

- β_1 captures the permanent differences between control and treatment group

- δ_1 captures the effect of the treatment (in second year and only for treatment group)

- It can be proven that the estimator of δ_1 is equal to:

$$\widehat{\delta_1} = (\overline{y}_{B.2} - \overline{y}_{B.1}) - (\overline{y}_{A.2} - \overline{y}_{A.1})$$

- Interpretation:

  - effect of the treatment is calculated as the change in y observed for the treatment group.
  - effect of the year is controlled for by subtracting the same change calculated for control group.

- Hence the name difference in differences (DID) estimator

- It is possible make this model more complicated by taking into account additional regressors

- Requirement for consistency: treatment not related to factors that affect y and are not observed

Example 1. *(Wooldridge: Mayer, Viscusi, Durbin 1995) Length of time on time on workers compensation*

1980 Kentucky rises the cap on the weekly earning covered by workers' compensation. This change does not affect low-wage workers (below the old cap). Control group: low-wage workers. Treatment group: high-wage workers. Question: what is the effect of compensation on the duration of stay out of work. Regression:

$$\log(\textit{durat}) = \underset{(0.031)}{1.126} + \underset{(.0447)}{.0077}\,\textit{afchnge} + \underset{(.047)}{.256}\,\textit{highearn} + \underset{(.069)}{.\ 191}\,\textit{afchnge} \cdot \textit{highearn}$$

$$N = 5626 \qquad R^2 = .021$$

Result: average duration of the stay increased by $\widehat{\delta}_1 = 19\%$ *due to higher cap. No tendency to longer stay for both groups: coefficient of afchnge insignificant. High earners has a tendency to stay on compensation about* $\left(\exp\left(\widehat{\beta}_1\right) - 1\right) = 29.2\%$ *longer than low earners.*

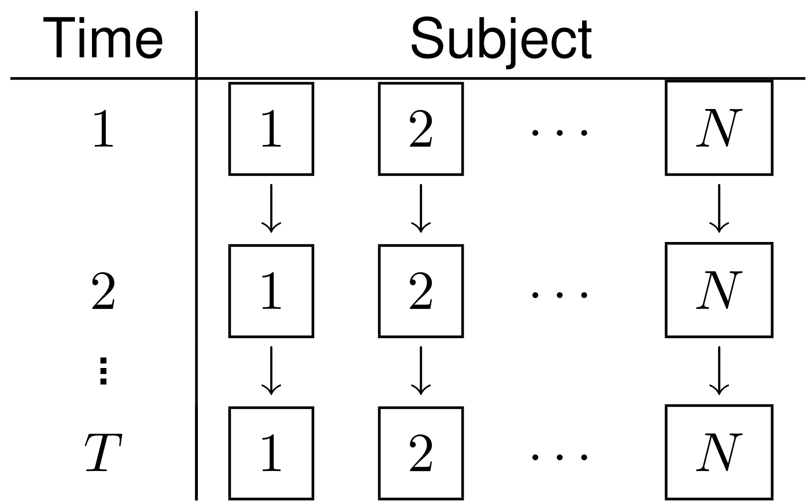
Other sampling schemes

- **Stratified samples**
- Number of strata (subgroups) small, number of observations in the strata large
- Problems: coefficients can be different for each strata, random error correlated inside strata (unobserved strata effect)
- **Spatial dependence**
- No problem if the dependence is only between regressors
- Problem if the correlation exists between error terms in neighboring units

- Difficult to derive the asymptotic properties of estimators
- **Clustered samples**
- Large number of clusters, number of observations inside clusters small
- Example: error terms within family can be correlated
- Within cluster correlation between error terms is not difficult to deal with
- OLS estimators are still unbiased but variance matrix need to be adjusted

Panel data

- Panel data is the data collected for the same group of subjects for several years



where $\boxed{i}$ is a subject number i the same for all times the data is collected

- Panel data is different from cross sections over time as it is collected for the same group for several years

Linear Unobserved effects Panel Data Models

- Unobserved Effects Model (index i for unit, t for time):

$$\mathrm{E}\left(y_{it} \middle| \boldsymbol{x}_t, c\right) = \boldsymbol{x}_{it}\, \boldsymbol{\beta} + c_i$$

- c_i is an unobserved and time constant
- c_i is called *unobserved (individual) effect (component)* also *unobserved (individual) heterogeneity*
- u_{it} is called *idiosyncratic error (disturbance)*
- Model in error form

$$y_{it} = \boldsymbol{x}_{it}\, \boldsymbol{\beta} + c_i + u_{it}$$

and $\mathrm{E}(u_{it}) = 0$

- **Balanced panel**
- All observations are available for all the units
- The unbalanced panel is a bit more difficult to deal with but necessary corrections are already available in same econometric packages
- **Asymptotic analysis**
- Usually in the panel the number of observation in cross section N is much larger than number of cross sections T
- It is why it is important to establish that the estimator is consistent for $N \to \infty$ even if T is constant

- If T goes to infinity when T stays constant (e.g. cross section of countries) different kind of arguments are needed

- **Advantages of using panel data models**

- Efficiency: panel data structure can be used for more efficient estimation by taking in account the correlation between unobserved individual effects (Random effect estimation)

- Consistency: if the individual effects are correlated with explanatory variables than the panel data structure can be used to obtain the consistent estimator using the fact that individual effects are constant over time (fixed effect estimation)

Example 2. *(Wooldridge) Program evolution. Model for evaluating the* *effects of training*

$$\log(wage_{it}) = \theta_t + \boldsymbol{z}_{it}\boldsymbol{\beta} + \boldsymbol{\delta}_1 \textit{prog}_{it} + c_i + u_{it}$$

θ_t *time varying intercept,* $\boldsymbol{z}_{it}$ *observable characteristics,* $\textit{prog}_{it}$ *dummy for participation.*

Takes into account individual effect c_i *(say ability). This effect can be correlated with participation if decision to participate are related to ability (self selection).*

Example 3. *(Wooldridge) Lagged Dependent Variable. Model of wage determination*

$$\log(\textit{wage}_{it}) = \beta_1 \log\left(\textit{wage}_{i,t-1}\right) + c_i + u_{it}$$

Our interest is in coefficient β_1 *related to wage persistence after controlling for individual heterogeneity.*

Estimating Unobserved Effects by $POLS$

- Model

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}$$

- $v_{it} = c_{it} + u_{it}$ are composite errors.

- Necessary condition for consistency: $\mathrm{E}\left(\boldsymbol{x}_{it}' v_{it}\right) = 0$ which implies that $\mathrm{E}\left(\boldsymbol{x}_{it} c_i\right) = 0$

- Individual effects and exogenous variables uncorrelated

- In this case Pooled OLS is consistent

- But: v_{it} are serially correlated due to the presence of c_i in period $t = 1, 2, \ldots$

- The it is necessary to use robust variance matrix for inference

$$\widehat{\boldsymbol{V}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i \widehat{\boldsymbol{u}}_i \widehat{\boldsymbol{u}}_i \boldsymbol{X}_i\right) \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1}$$

- Where $\boldsymbol{X}_i$ are observation matrix for $t = 1, \ldots, T$ for unit i and $\widehat{\boldsymbol{u}}_i$ is vector of $POLS$ residuals for unit i.

- This variance matrix can be obtained in STATA by defining units as clusters.

Random effect method

- Assumptions for consistency Random effects model

  1. Strict exogeneity: $\mathrm{E}\left(u_{it} \mid \boldsymbol{x}_i, c_i\right) = 0$ for $t = 1, \ldots, T$ $\;(FGLS)$
  2. Independence between c_i and $\boldsymbol{x}_i$, $\mathrm{E}\left(c_i \mid \boldsymbol{x}_i\right) = 0$
     where $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})$

- Model

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}$$

- $v_{it} = c_i + u_{it}$ are composite errors.

- In matrix notation

$$\boldsymbol{y}_i = \boldsymbol{X}_i\,\boldsymbol{\beta} + \boldsymbol{v}_i$$

- Idiosyncratic errors u_{it} are homoscedastic and uncorrelated
  - $\mathrm{E}\left(u_{it}^2\right) = \sigma_u^2$
  - $\mathrm{E}\left(u_{it}u_{is}\right) = 0$
  - not correlated with individual effects $\mathrm{E}\left(u_{it}, c_i\right) = 0$

- Under assumptions made variance of composite error is equal to

$$\mathrm{E}\left(v_{it}^2\right) = \mathrm{E}\left(c_i^2\right) + 2\,\mathrm{E}\left(c_i u_{it}\right) + \mathrm{E}\left(u_{it}^2\right) = \sigma_c^2 + \sigma_u^2$$

  where σ_c^2 is the variance of c_i

- For $t \neq s$, the covariance between v_{it} and v_{is} is equal to

$$\mathrm{E}\left(v_{it}v_{is}\right) = \mathrm{E}\left[\left(c_i + u_{it}\right)\left(c_i + u_{is}\right)\right] = \mathrm{E}\left(c_i^2\right) = \sigma_c^2$$

- Then

$$\boldsymbol{\Omega} = \mathrm{E}\left(\boldsymbol{v}_i \boldsymbol{v}_i'\right) = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & & & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

- Assumptions for efficiency of Random effects: $\mathrm{E}\left(\boldsymbol{u}_i \boldsymbol{u}_i' \middle| \boldsymbol{x}_i\right) = \sigma_u^2 \boldsymbol{I}$ and $\mathrm{E}\left(c_i^2 \middle| \boldsymbol{x}_i\right) = \sigma_c^2$

- Generalised Least Squares (GLS) - general case

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}$$

where $\overline{\boldsymbol{\Omega}} = \mathrm{E}\left(\boldsymbol{v}\boldsymbol{v}'\right)$, where $\boldsymbol{u} = \left(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T\right)$

- But for panels only v_{it} and v_{is} are correlated for $t \neq s$, v_{it} and v_{js} are uncorrelated for $i \neq j$

- Then $\overline{\mathbf{\Omega}}$ has block diagonal form

$$\overline{\mathbf{\Omega}} = \begin{bmatrix} \mathbf{\Omega} & & 0 \\ & \ddots & \\ 0 & & \mathbf{\Omega} \end{bmatrix}$$

and GLS estimator have the special form

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

- $\mathbf{\Omega}$ is not known but in only depends on 2 parameters: σ_u^2 and σ_c^2

- Assume that we have consistent estimators of σ_u^2 and σ_c^2

- Random effects estimator ($FGLS$)

$$\widehat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{y}_i \right)$$

- To make this formula operational we have to find estimators for σ_u^2 and σ_c^2

- Denote variance of composite error as $\sigma_v^2 = \sigma_u^2 + \sigma_c^2$

- As $POLS$ estimator of $\boldsymbol{\beta}$ is consistent for RE assumptions then the $POLS$ estimator $\widehat{\sigma}_v^2$ is also a consistent estimator of σ_v^2.

- Covariance between v_{it} and v_{is} is equal to σ_c^2.

- Empirical covariance between the residuals of $POLS$ can be used then as estimator of σ_c^2:

$$\widehat{\sigma}_c^2 = \frac{1}{[NT\,(T-1)/\,2-K]} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \widehat{\widehat{v}}_{it} \widehat{\widehat{v}}_{is}$$

- As $\sigma_v^2 = \sigma_c^2 + \sigma_u^2$, the estimator of σ_u^2 can be calculated as $\widehat{\sigma}_u^2 = \widehat{\sigma}_v^2 - \widehat{\sigma}_c^2$ (can be negative)

Example 4. *Wooldridge (RE estimation of the effects of Job Training Grants). Estimate the effects of job training grants on firms scrap rates.* 54 *firms, reported scrap rates for* $1987, 1988, 1989$*. Grants not awarded in* 1987*. No firm can receive grants twice. Effect of the grant can persist. Problem:*

scrap rates can be related to unobserved individual characteristics.

$$\log(scrap) = \underset{(.241)}{.415} + \underset{(.109)}{.093}\, d88 - \underset{(.132)}{.270}\, d89 + \underset{(.411)}{.548}\, union$$
$$- \underset{(.148)}{.215}\, grant - \underset{(.205)}{.377}\, grant_{-1}$$

- Under assumption $\mathrm{E}(\boldsymbol{u}_i \boldsymbol{u}_i' | \boldsymbol{x}_i, \boldsymbol{c}_i) = \sigma_u^2 \boldsymbol{I}_T$ and $\mathrm{E}\left(c_i^2 \middle| \boldsymbol{x}_i\right) = \sigma_c^2$ the random effect estimator is the most efficient

- If there is a heteroscedasticity (unequal variances for units), or the variance matrix has no random effect structure we can use robust variance estimator.

Fixed effects methods

- Model
$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

- We assume that individual effects c_i are correlated with $\boldsymbol{x}_{it}$

- For consistency of fixed effects estimator we only need strict exogeneity assumption
$$\mathrm{E}\left(u_{it} | \boldsymbol{x}_i, c_i\right) = 0 \text{ for } t = 1, \ldots, T$$

- Because the assumption $\mathrm{E}\left(c_i | \boldsymbol{x}_i\right) = 0$ is not needed the fixed effect analysis is more robust then random effect analysis

- **Problem**: it is not possible to analyze with fixed effects methods the influence of time constant variable

- For arbitrary correlation between c_i and time constant variable x_k there is no way to distinguish the effects of these two variables.

Example 5. *It is not possible to analyze the influence of gender on wages with fixed effect estimator as gender does not changes for individuals.*

- Fixed effect transformation (within transformation) is used to eliminate individual effect c_i

- We average equation $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$ over time and get:

$$\overline{y}_i = \overline{\boldsymbol{x}}_i\boldsymbol{\beta} + c_i + \overline{u}_i$$

where $\overline{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$, $\overline{\boldsymbol{x}}_i = \frac{\sum_{t=1}^{T} \boldsymbol{x}_{it}}{T}$, $\overline{u}_i = \frac{\sum_{t=1}^{T} u_{it}}{T}$

- Subtracting this equation from original model we obtain FE transformed equation

$$y_{it} - \overline{y}_i = \left(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i\right)\boldsymbol{\beta} + \left(u_{it} - \overline{u}_i\right)$$

or

$$\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{it}\,\boldsymbol{\beta} + \ddot{u}_{it}$$

- This equation can be consistently estimated by $POLS$ as

$$\mathrm{E}\left(u_{it}|\,\boldsymbol{x}_i, c_i\right) = 0 \Longrightarrow \mathrm{E}\left(\overline{u}_{it}|\,\boldsymbol{x}_i, c_i\right) = 0$$

and so

$$\mathrm{E}\left(\ddot{u}_{it}|\,\boldsymbol{x}_i, c_i\right) = 0 \Longrightarrow \mathrm{E}\left(\ddot{u}_{it}|\,\ddot{\boldsymbol{x}}_{it}, c_i\right) = 0$$

which is sufficient for consistency of $POLS$

- The Fixed Effect (FE) estimator is thus calculated as $POLS$ estimator of regression of $\ddot{y}_{it}$ on $\ddot{\boldsymbol{x}}_{it}$

- The correct estimate of σ_u^2 for this estimator (unbiased conditional on $\boldsymbol{X}$) is $\sigma_u^2 = \frac{SSR}{[N(T-1)-K]}$

- This estimator is also called within estimator as it only uses the variation within units (not between units)

- The equivalent form of this estimator can be calculated running $POLS$ regression of y_{it} on $\boldsymbol{x}_{it}$, $d1_i, d2_i, \ldots, dN_i$ where dn_i is the dummy variable for unit i so that $dn_i = 1$ if $n = i$, $dn_i = 0$ if $n \neq i$

- Coefficients estimated for $d1_i, \ldots, dN_i$ are estimates of individual effects c_i. This estimates are unbiased but inconsistent unless $T \to \infty$

- This form of FE estimator is sometimes called least squares dummy variable estimator ($LSDV$)

- To estimate $\widehat{c}_i$ without using $LSDV$ (which can be infeasible because of large number of parameters) we can use the standard form of FE estimator to obtain $\widehat{\boldsymbol{\beta}}_{FE}$ and then calculate

$$\widehat{c}_i = \overline{y}_i - \overline{\boldsymbol{x}}_i \widehat{\boldsymbol{\beta}}_{FE}$$

- Under assumption $\mathrm{E}\left(\boldsymbol{u}_i \boldsymbol{u}_i' \middle| \boldsymbol{x}_i, \boldsymbol{c}_i\right) = \sigma_u^2 \boldsymbol{I}_T$ fixed effect estimator is the most efficient

- Variance of $\ddot{u}_{it}$ is equal to

$$\begin{aligned} \mathrm{E}\left(\ddot{u}_{it}^2\right) &= \mathrm{E}\left(u_{it}^2\right) + \mathrm{E}\left(\overline{u}_i^2\right) - 2\,\mathrm{E}\left(u_{it}\overline{u}_i\right) \\ &= \sigma_u^2 + \frac{\sigma_u^2}{T} - 2\frac{\sigma_u^2}{T} = \sigma_u^2\left(1 - \frac{1}{T}\right) \end{aligned}$$

- Transformed errors are homoscedastic

- Covariance between $\ddot{u}_{it}$ and $\ddot{u}_{is}$ is equal to

$$\begin{aligned}\mathrm{E}\left(\ddot{u}_{it}\ddot{u}_{is}\right) &= \mathrm{E}\left[\left(u_{it}-\overline{u}_i\right)\left(u_{is}-\overline{u}_i\right)\right] \\ &= \mathrm{E}\left(u_{it}u_{is}\right)-\mathrm{E}\left(u_{it}\overline{u}_i\right)-\mathrm{E}\left(u_{is}\overline{u}_i\right)+\mathrm{E}\left(\overline{u}_i^2\right) \\ &= 0-\frac{\sigma_u^2}{T}-\frac{\sigma_u^2}{T}+\frac{\sigma_u^2}{T}=-\frac{\sigma_u^2}{T}\end{aligned}$$

- Transformed errors are autocorrelated with correlation coefficient equal to $Corr\left(\ddot{u}_{it},\ddot{u}_{is}\right)=-\frac{1}{1-T}$

- Care is needed when testing for autocorrelation (tests are not possible $T=2$ and for $T>2$ have special form)

- If the problem of serial correlation of u_{it} or heteroscedasticity of u_{it} is detected we can use $FEGLS$ (Fixed Effect GLS):

$$\left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \widehat{\boldsymbol{\Omega}}^{-1} \ddot{\boldsymbol{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \ddot{\boldsymbol{X}}_i' \widehat{\boldsymbol{\Omega}}^{-1} \ddot{\boldsymbol{y}}_i\right)$$

where $\widehat{\boldsymbol{\Omega}} = N^{-1} \sum_{i=1}^{N} \widehat{\widehat{\boldsymbol{u}}}_i \widehat{\widehat{\boldsymbol{u}}}_i'$ and $\boldsymbol{u}_i$ are calculated from FE regression

Example 6. *(Wooldridge) FE estimation of the effects of the Job training Grants. We have to drop variable unions because it does not change in time for units in the sample*

$$\log(scrap) = -\underset{(.109)}{.080}\,\textit{d88} - \underset{(.133)}{.247}\,\textit{d89} - \underset{(.151)}{.252}\,\textit{grant} - \underset{(.210)}{.422}\,\textit{grant}_{-1}$$

Compared with random effects grants has larger effect on scrap rate.

First differencing Methods

- Assumption needed for consistency - the same as for fixed effect

$$\mathrm{E}\left(u_{it} \middle| \boldsymbol{x}_i, c_i\right) = 0 \text{ for } t = 1, \ldots, T$$

- Differencing transformation is used to eliminate individual effects c_i
- Model

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

- We take first differences of both sides

$$\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$$

- The first observation ($T = 1, i = 1, \ldots, N$) has to be omitted
- Only the time varying variables can be included, time constant variables are transformed to zero
- We obtain first difference (FD) estimator by running $POLS$ of Δy_{it} on $\Delta \boldsymbol{x}_{it}$
- This estimator is consistent as $\mathrm{E}\left(\Delta \boldsymbol{x}_{it}' \Delta u_{it}\right) = 0$
- Under assumption $\mathrm{E}\left(\Delta \boldsymbol{u}_i \Delta \boldsymbol{u}_i' | \, \boldsymbol{x}_i, \boldsymbol{c}_i\right) = \sigma_u^2 \boldsymbol{I}_T$ the fixed effect estimator is the most efficient

Example 7. *Wooldridge) FE estimation of the effects of the Job training Grants. We have to drop variable unions because it does not change in time for units in the sample*

$$\log\left(\Delta scrap\right) = -\underset{(.091)}{.091}\, \textit{d88} - \underset{(.125)}{.096}\, \textit{d89} - \underset{(.131)}{.233}\, \Delta \textit{grant} - \underset{(.235)}{.351}\, \Delta \textit{grant}_{-1}$$

Differences are not great for the FE and FD estimators

Comparison of estimators

- Random effects and fixed effects
- RE estimator consistent if $\boldsymbol{c}_i$ and $\boldsymbol{x}_i$ are uncorrelated and efficient if $\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'\middle|\,\boldsymbol{x}_i,\boldsymbol{c}_i\right)=\sigma_u^2\boldsymbol{I}_T$ and $\mathrm{E}\left(c_i^2\middle|\,\boldsymbol{x}_i\right)=\sigma_c^2$
- FE estimator consistent even if $\boldsymbol{c}_i$ and $\boldsymbol{x}_i$ are correlated and is efficient if $\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'\middle|\,\boldsymbol{x}_i,\boldsymbol{c}_i\right)=\sigma_u^2\boldsymbol{I}_T$
- FD estimator consistent even if $\boldsymbol{c}_i$ and $\boldsymbol{x}_i$ are correlated and is efficient if $\mathrm{E}\left(\Delta\boldsymbol{u}_i\Delta\boldsymbol{u}_i'\middle|\,\boldsymbol{x}_i,\boldsymbol{c}_i\right)=\sigma_u^2\boldsymbol{I}_T$
- Significant differences between RE and FE, FD estimates indicate that assumption $\mathrm{E}\left(\boldsymbol{x}_i'c_i\right)=0$ is false

- To test whether the difference use the Hausmann test
- It is possible to prove that random effect estimator can be calculated by running $POLS$ on transformed equation

$$y_{it} - \lambda \overline{y}_i = (\boldsymbol{x}_{it} - \lambda \overline{\boldsymbol{x}}_i) + v_{it} - \lambda \overline{v}_i$$

where $\lambda = 1 - \left[\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_c^2}\right]^{\frac{1}{2}}$

- This transformation is called quasi-time demeaning
- Special cases:
  - $T \to \infty$ or $\frac{\sigma_c^2}{\sigma_u^2} \to \infty$ RE and FE give the same results
  - $\lambda = 0$ (never exactly true but close to for σ_u^2 large relative to σ_c^2) we obtain $POLS$

Example 8. *(Wooldridge) Job Training Grants* $\widehat{\lambda} \approx .797$*. This explains why the estimated coefficient similar.*

Panel models under sequential moment restrictions

- Normally to obtain consistency of RE, FE, FD estimators we assume strict exogeneity:

$$\mathrm{E}\left(u_{it} | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}\right)$$

- Idiosyncratic error is independent on past, contemporaneous and *future* values of explanatory variables $\boldsymbol{x}_i$

- Now we assume that in the model

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad \text{for } t = 1, 2, \ldots, T$$

and we assume that $\boldsymbol{x}_{it}$ can be correlated with c_i and u_{it} can be correlated with *future* values of $\boldsymbol{x}_{it}$

- However we introduce the *sequential moment restriction* of the form

$$\mathrm{E}\left(u_{it} \middle| \boldsymbol{x}_{it}, \boldsymbol{x}_{it-1}, \ldots, \boldsymbol{x}_{i1}\right) = 0$$

- That is u_{it} cannot be correlated with present and past values of $\boldsymbol{x}_i$

- We say that $\boldsymbol{x}_{it}$ is sequentially exogenous conditional on the unobserved effect

- Conditional expectation form of the model

$$\mathrm{E}\left(y_{it} \middle| \boldsymbol{x}_{it}, \boldsymbol{x}_{it-1}, \ldots, \boldsymbol{x}_{i1}, c_i\right) = \mathrm{E}\left(y_{it} \middle| \boldsymbol{x}_{it}, c_i\right) = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i$$

Example 9. *(Wooldrodge) Dynamic unobserved effects model)*

$$y_{it} = \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho_1 y_{i,t-1} + c_i + u_{it}$$

and so $\boldsymbol{x}_{it} \equiv (\boldsymbol{z}_{it}, y_{i,t-1})$. *Therefore* $(\boldsymbol{x}_{it}, \boldsymbol{x}_{it-1}, \ldots, \boldsymbol{x}_{i1}) = (\boldsymbol{z}_{it}, y_{i,t-1}, \ldots, \boldsymbol{z}_{i1}, y_{i,0})$ *and sequential moment restriction requires*

$$\begin{aligned} \mathrm{E}\left(y_{it} \middle| \boldsymbol{z}_{it}, y_{i,t-1}, \ldots, \boldsymbol{z}_{i1}, y_{i,0}, c_i\right) &= \mathrm{E}\left(y_{it} \middle| \boldsymbol{z}_{it}, y_{i,t-1}, c_i\right) \\ &= \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho_1 y_{i,t-1} + c_i \end{aligned}$$

- If sequential moment condition are true but strict exogeneity is false RE, FE, FD estimators are inconsistent. For example

$$\operatorname{plim}\left(\widehat{\boldsymbol{\beta}}_{FE}\right) = \boldsymbol{\beta} + \left[T^{-1}\sum_{i=1}^{T}\mathrm{E}\left(\ddot{\boldsymbol{x}}_{it}\ddot{\boldsymbol{x}}_{it}'\right)\right]^{-1}\left[T^{-1}\sum_{i=1}^{T}\mathrm{E}\left(\ddot{\boldsymbol{x}}_{it}' u_{it}\right)\right]$$

but $\mathrm{E}\left(\ddot{\boldsymbol{x}}_{it}' u_{it}\right) = \mathrm{E}\left[\left(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{it}\right) u_{it}\right] = -\mathrm{E}\left(\overline{\boldsymbol{x}}_{it} u_{it}\right) = O\left(T^{-1}\right) \neq 0$.

- Asymptotic bias is of order T^{-1} but usually for panels T is small

- If $\boldsymbol{x}_{it}$ is stationary the FE estimator is better than FD estimator as this bias is of order T^{-1} for FE but does not depend on T for FD

- However it is possible to derive estimators which are consistent under sequential moment conditions

- If we take first differences of the initial model we get

$$\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad \text{for } t = 2, \ldots, T$$

- Under assumptions made

$$\mathrm{E}\left(\boldsymbol{x}'_{is}u_{it}\right) = 0, \quad \text{for } s = 1, 2, \ldots, t$$

- and

$$\mathrm{E}\left(\Delta \boldsymbol{x}'_{it}\Delta u_{it}\right) \neq 0$$

as $\boldsymbol{x}_{it-1}$ can be correlated with u_{it}.

- But

$$\mathrm{E}\left(\boldsymbol{x}_{is}'\Delta u_{it}\right) = 0, \quad \text{for } s = 1, 2, \ldots, t-1$$

and

$$\mathrm{E}\left(\Delta\boldsymbol{x}_{is}'\Delta u_{it}\right) = 0, \quad \text{for } s = 1, 2, \ldots, t-1$$

- Then the variables $\boldsymbol{x}_{it}^{o} = \boldsymbol{x}_{it-1}, \ldots, \boldsymbol{x}_{i,0}$ or $\boldsymbol{x}_{it}^{o} = \Delta\boldsymbol{x}_{it-1}, \ldots, \Delta\boldsymbol{x}_{i,0}$ (or any linear combination of them) can be used as instrumental variables in differenced equation

- So the first differenced equation can be estimated with $2SLS$

- It make sense to assume that rank condition $\mathrm{E}\left(\Delta\boldsymbol{x}_{it}\Delta\boldsymbol{x}_{it-1}'\right) = K$ is true. This choice of instruments is only possible if $T \geq 3$

- For $T = 2$ we can use $\boldsymbol{x}_{it-1}$ but often correlation between $\Delta\boldsymbol{x}_{it}$ and $\boldsymbol{x}_{it-1}$ is small

- The efficient estimator in this context is the GMM estimator making use of all moment restrictions. However, as this estimator is using a lot of overidentifing restrictions the small sample behavior can be poor

Example 10. *(Wooldridge) Testing for persistence of crime rate. Crime rate can be related to unobserved county effect. As the equation includes lagged variable $y_{i,t-1}$ we use FD model and $y_{i,t-1}$, $y_{i,t-2}$ are used as instruments for Δy_{it}. Relation between Δy_{it} and $y_{i,t-1}$, $y_{i,t-2}$ significant p-value=0.023 (rank condition probably true)*

$$\Delta \log(\textit{crmrte}) = \underset{(.040)}{.065} + \underset{(.497)}{.212}\, \Delta \log(\textit{crmrte})_{-1}$$

we can not reject that $H_0 : \rho_1 = 0$ (crime rate not persistent)

- If the u_{it} are uncorrelated in original model than they are correlated in differenced model. This problem can be solved by using robust variance matrix.

- It can happen that we have model

$$y_{it} = \boldsymbol{z}_{it}\gamma + \boldsymbol{w}_{it}\boldsymbol{\delta} + c_i + u_{it}, \quad \text{for } t = 1, 2, \ldots, T$$

where $\boldsymbol{z}_{it}$ are strictly exogenous but $\boldsymbol{w}_{it}$ are sequentially exogenous.

- In this case we can estimate with pooled $2SLS$ equation

$$\Delta y_{it} = \Delta \boldsymbol{z}_{it}\gamma + \boldsymbol{\Delta w}_{it}\boldsymbol{\delta} + \Delta u_{it}, \quad \text{for } t = 2, \ldots, T$$

with the use of instruments $\boldsymbol{z}_{it}, \boldsymbol{x}_{it-1}, \ldots, \boldsymbol{x}_{i,0}$ or any linear combination of them.

- Typical application of this method is the model

$$y_{it} = \boldsymbol{z}_{it}\,\boldsymbol{\gamma} + \rho_1 y_{it-1} + c_i + u_{it}$$

- Sometimes we have contemporaneous correlation between explanatory variables and idiosyncratic error

- Then $\mathrm{E}\left(\boldsymbol{z}_{is}' u_{it}\right) = 0$ for all s, t but we allow w_{it} to be contemporaneously correlated with u_{it}

Example 11. *(Wooldridge) Effects of smoking on hourly wage*

$$\log\left(\textit{wage}_{it}\right) = \boldsymbol{z}_{it}\boldsymbol{\gamma} + \delta_1 \textit{cigs}_t + c_i + u_{it}$$

Cigarette smoking depends on individual characteristics. On the other hand wage can affect smoking as well (simultaneity).

- In this situation we usually take first defiances and estimate the equation with the use of instruments from outside of the model (although we could use $\boldsymbol{z}_i'$ and $\boldsymbol{w}_{it-2}, \ldots, \boldsymbol{w}_{i0}$ as instruments too).

Exercise 12. *cont. For the cigs$_t$ the valid instrument could be the local prices of cigarettes. Determines the consumption of cigarettes but is not related to wage.*

- If we have no lagged variables in the system we could also use fixed effect transformation followed by $2SLS$.

Models with individual-specific slopes

- The simplest model to consider is *random trend model*

$$y_{it} = c_i + g_i t + \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}$$

- The growth rates (in loglinear model) are different for individuals

- Strict exogeneity assumption and conditional mean specification

$$\mathrm{E}\left(u_{it} \middle| \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, c_i, g_i\right) = 0$$

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

- One approach is to first difference the original equation

$$\Delta y_{it} = g_i + \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \text{ for } t = 2, \ldots, T$$

  and estimate this equation with FE or FD estimator

- To apply first differences we need $T \geq 2$ and to apply FE or FD estimators afterwards we need $T \geq 3$

- General model with individual-specific slopes

$$y_{it} = \boldsymbol{z}_{it}\boldsymbol{a}_i + \boldsymbol{x}_{it}\,\boldsymbol{\beta} + u_{it}$$

- Strict exogeneity assumption

$$\mathrm{E}\left(u_{it} | \, \boldsymbol{z}_i, \boldsymbol{x}_i, \boldsymbol{a}_i\right) = 0 \text{ for } t = 1, \ldots, T$$

- We can write the model in matrix form

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{a}_i + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{u}_i$$

- Define matrix $\boldsymbol{M}_i = \boldsymbol{I}_T - \boldsymbol{Z}_i \left(\boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1} \boldsymbol{Z}_i'$ and multiply the equation of interest

$$\boldsymbol{M}_i \boldsymbol{y}_i = \underbrace{\boldsymbol{M}_i \boldsymbol{Z}_i}_{0} \boldsymbol{a}_i + \boldsymbol{M}_i \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{M}_i \boldsymbol{u}_i$$

$$\ddot{\boldsymbol{y}}_i = \ddot{\boldsymbol{X}}_i \boldsymbol{\beta} + \boldsymbol{M}_i \boldsymbol{u}_i$$

where $\ddot{\boldsymbol{y}}_i, \ddot{\boldsymbol{X}}_i$ are residuals from the regression of $\boldsymbol{y}_i, \boldsymbol{X}_i$ on $\boldsymbol{Z}_i$

- Under strict exogeneity assumption $\mathrm{E}\left(\ddot{\boldsymbol{X}}_i' \boldsymbol{u}_i\right) = 0$ and Pooled OLS applied to transformed equation is consistent

- The rank condition $\mathrm{E}\left(\ddot{\boldsymbol{X}}_i' \ddot{\boldsymbol{X}}_i\right) = K$ will fail if there are elements of $\boldsymbol{x}_{it}$ which not vary with time

- It is also possible to obtain consistent estimate of $\boldsymbol{\alpha} = \mathrm{E}(\boldsymbol{a}_i)$ as

$$\alpha = \mathrm{E}\left[\left(\boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\,\boldsymbol{\beta}\right)\right]$$

  then the estimate is

$$\widehat{\alpha} = \sum_{i=1}^{N}\left[\left(\boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\,\widehat{\boldsymbol{\beta}}_{FE}\right)\right]$$

- Unbiased but inconsistent estimate of individual slopes is given by

$$\widehat{\boldsymbol{a}}_i = \left(\boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\,\widehat{\boldsymbol{\beta}}_{FE}\right)$$

Hausman and Taylor type estimators

- Common problem in estimating panel model with fixed effect - not possible to take into account time invariant variables
- Model

$$y_{it} = \boldsymbol{z}_i \boldsymbol{\gamma} + \boldsymbol{x}_{it} \boldsymbol{\beta} + c_i + u_{it}$$

- Strict exogeneity assumption

$$\mathrm{E}\left(u_{it} \middle| \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, c_i \right) = 0$$

- The FE and FD estimators eliminate γ

- But if $\mathrm{E}\left(\boldsymbol{z}_i'\boldsymbol{c}_i\right) = 0$ we can estimate γ using the fact that $\mathrm{E}\left(\boldsymbol{z}_i'\boldsymbol{z}_i\right)\boldsymbol{\gamma} = \mathrm{E}\left[\boldsymbol{z}_i'\left(\overline{y}_i - \overline{\boldsymbol{x}}_i\boldsymbol{\beta}\right)\right]$

$$\gamma = \left[N^{-1}\sum_{i=1}^{N}\boldsymbol{z}_i\boldsymbol{z}_i'\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{z}_i'\left(\overline{y}_i - \overline{\boldsymbol{x}}_i\widehat{\boldsymbol{\beta}}_{FE}\right)\right]$$

- General setup: $\boldsymbol{z}_i = \left(\boldsymbol{z}_{i,1}, \boldsymbol{z}_{i,2}\right)$, $\boldsymbol{x}_{it} = \left(\boldsymbol{x}_{it,1}, \boldsymbol{x}_{it,2}\right)$ and $\boldsymbol{z}_{i1}$, $\boldsymbol{x}_{it,1}$ and uncorrelated with c_i.

- Necessary condition $TK_1 \geq J_2$

- where K_1 number of variables in x_{i1t}, J_2 number of variables in z_{i2}

- Hausman-Taylor estimator:

1. Perform pooled $2SLS$ using IV $(\boldsymbol{z}_{i,1}, \ddot{\boldsymbol{x}}_{it,1}, \boldsymbol{x}^o_{it,1})$
   where $\boldsymbol{x}^o_{it1} = \left(\boldsymbol{x}^o_{it,1}, \boldsymbol{x}^o_{it-1,1}, \ldots, \boldsymbol{x}^o_{i1,1}\right)$
2. Calculate $\widehat{\sigma}^2_c, \widehat{\sigma}^2_u$ and $\widehat{\lambda}$
3. Quasi demean dependent, independent and instrumental variables
4. Calculate $2SLS$ with quasi demeaned variables