5 Panel Data

- panel data include multiple draws on the same basic unit of observation ('group')
- typically, multiple draws over time, but panel data need not require a time dimension
- examples
  - individuals, firms, or countries observed at multiple time periods
  - multiple individuals within a household observed at a point in time
  - multiple employees within a firm observed at a point in time
- can have 3 (or higher) dimensional panels (e.g., multiple individuals within a household observed at multiple points in time)

- data structure

$$\{y_{it}, x_{it}\}_{i=1,\ldots,N;t=1,\ldots,T}$$

where total sample size $= NT$

  - examples

    * i indexes individuals, firms, or countries; t indexes time periods
    * i indexes individuals; t indexes households
    * i indexes employees; t indexes firms

- in microeconometric studies, typically N is large, T is small; macroeconometrics may be the reverse
- asymptotics can be performed on $N \longrightarrow \infty, T \longrightarrow \infty$, or both

5.1 Pooled OLS

- model

$$y_{it} = \alpha + x_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

- estimation via OLS
- identical to usual OLS, only now sample size is NT
- usual assumptions required for unbiasedness, consistency, etc.
- extensions
  - time trend
    - linear time trend

      $$y_{it} = \alpha + \lambda t + x_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

      which allows the intercept to trend linearly over time, changing by λ each period
    - quadratic time trend

      $$y_{it} = \alpha + \lambda_1 t + \lambda_2 t^2 + x_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

      which allows the intercept to follow a more general time trend
  - structural break

    $$y_{it} = \begin{cases} \alpha_1 + x_{it}\beta_1 + \varepsilon_{it} & \text{if } t \leqslant \widetilde{T} \\ \alpha_2 + x_{it}\beta_2 + \varepsilon_{it} & \text{if } t > \widetilde{T} \end{cases}$$

    where $\widetilde{T}$ is the date of the structural break

∗ could have multiple breaks if panel is long enough

∗ Chow test

$$H_o : \alpha_1 = \alpha_2, \beta_1 = \beta_2$$
$$H_1 : \text{not all equal}$$

has a test statistic of

$$F_{K+1,NT-2K-2} = \frac{(SSR_R - SSR_1 - SSR_2)/(K+1)}{(SSR_1 + SSR_2)/(NT - 2K - 2)}$$

where

· SSR_R = SSR from pooled (restricted) model

· SSR_1 = SSR from OLS using only obs with $t \leqslant \widetilde{T}$

· SSR_2 = SSR from OLS using only obs with $t > \widetilde{T}$

· K = # of x's

∗ alternative

· define

$$I_{it} = \begin{cases} 1 & \text{if } t > \widetilde{T} \\ 0 & \text{if } t \leqslant \widetilde{T} \end{cases}$$

· estimate via OLS

$$y_{it} = \alpha_1 + \widetilde{\alpha}_2 I_{it} + x_{it}\beta_1 + x_{it} I_{it} \widetilde{\beta}_2 + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

and test

$$\begin{aligned} H_o &: \widetilde{\alpha}_2, \widetilde{\beta}_2 = 0 \\ H_1 &: \text{not all} = 0 \end{aligned}$$

– time-specific intercepts

$$y_{it} = \alpha + \sum\nolimits_{s=2}^{T} \lambda_s D_{st} + x_{it}\beta + \epsilon_{it}$$

where

$$D_{st} = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$$

* equivalent to

$$y_{it} = \sum\nolimits_{s=1}^{T} \lambda_s D_{st} + x_{it}\beta + \varepsilon_{it}$$

where constant is no omitted to avoid perfect multicollinearity

* λ's capture effects of all variables that do not vary across i at a point in time
* any x's that do not vary across individuals are subsumed by λ's even if they vary over time
* more general than time trend since intercepts can potentially bounce all over

* Chow test

$$H_o \ : \ \lambda_1 = \cdots = \lambda_T$$
$$H_1 \ : \ \text{not all equal}$$

has a test statistic of

$$F_{T-1,NT-K-T} = \frac{(SSR_R - SSR_U)/(T-1)}{SSR_U/(NT-K-T)}$$

where

· SSR_R = SSR from restricted model (single intercept)

· SSR_U = SSR from unrestricted model (T intercepts)

· $K = \#$ of x's

* can interact time dummies with x's to allow β's to vary over time

* inclusion of all time dummies, and all interactions between time dummies and x's equivalent to OLS period-by-period

– difference-in-difference estimation

  * frequently used in policy analyses
  * examples
    · What was the impact of NJ's minimum wage hike?
    · What is the impact of legalized abortion on crime?
    · What is the impact of the death penalty on crime?
  * cross-sectional model

$$y_i = \alpha + x_i\beta + \delta D_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, ..., N$$

    where, say,

    · y = unemployment rate
    · x = macroeconomic variables
    · $D = 1$ if NJ (high MW), 0 for all other states (low MW)
  * potential shortcoming: what if there are *unobservable* differences between observations with the policy, and those without the policy
    · e.g., if NJ is different from other states for reasons not included in x, then $\text{Cov}(D, \varepsilon) \neq 0$
    · $\Longrightarrow \widehat{\delta}_{OLS}$ (and perhaps $\widehat{\beta}_{OLS}$) will be biased
  * panel data offers a potential solution
  * involves collecting data *prior to* policy implementation

∗ intuition

· cross-sectional model identifies δ by comparing the *level* of y in states with the policy to the *level* of y in states without the policy

· difference-in-difference model identifies δ by comparing the *change* in y in states from before and after the policy to the *change* in y in states with no policy change

∗ panel model

$$y_{it} = \alpha + x_{it}\beta + \lambda_1 D_i + \lambda_2 D2_t + \delta D_i D2_t + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma^2)$$

where, say,

· y = unemployment rate

· x = macroeconomic variables

· $D_i = 1$ if NJ ('policy changer'), 0 for all other states (no policy change)

· $D2_t = 1$ for periods in which NJ has a high MW, 0 for previous time periods

· $D_i D2_t = 1$ if NJ after policy change, 0 for all other observations

$*$ interpretation of parameters in the panel model (ignoring x's)

	pre-policy change $t = 1$	post-policy change $t = 2$
no policy change $D = 0$	α	$\alpha + \lambda_2$
'policy changer' $D = 1$	$\alpha + \lambda_1$	$\alpha + \lambda_1 + \lambda_2 + \delta$

which implies

· $\lambda_2 =$ difference (or change) over time in states without policy change ($\alpha + \lambda_2 - \alpha = \lambda_2$)

· $\lambda_2 + \delta =$ difference (or change) over time in states with policy change ($\alpha + \lambda_1 + \lambda_2 + \delta - (\alpha + \lambda_1) = \lambda_2 + \delta$)

· $\delta =$ difference in the the two differences ($\lambda_2 + \delta - \lambda_2 = \delta$), which is the additional *change* in states with the policy change

· $\widehat{\delta}_{POLS}$ known as DID estimator

* notes

  - λ_1 captures time-invariant differences in states with the policy change vs. states with no policy change; solves the omitted variable bias problem in cross-sectional models if the relevant omitted vars do not change over time
  - λ_2 captures changes over time that affect all states – policy changers and non-changers – equally
  - $\widehat{\delta}_{POLS}$ = unbiased estimate of policy impact if (i) λ_1 captures all differences between policy changers and non-changers, and (ii) the change in y over time (equal to λ_2) is idential for both policy changers and non-changers
  - if no x's, then

$$\widehat{\delta}_{POLS} = (\overline{y}_2^1 - \overline{y}_1^1) - (\overline{y}_2^0 - \overline{y}_1^0)$$

    where $\overline{y}_t^D$ = mean outcome in period t of states of type D

5.2 Fixed Effects

- motivation
  - OLS is biased if omitted vars are correlated with included x's
  - not always possible to find valid IVs
  - if omitted var does not vary over time (time invariant), panel data can yield estimates free from omitted variable bias
- same setup as before, but allow for individual-specific intercepts

$$y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

  - α_i = FE for group i (aka, unobserved effect, unobserved heterogeneity)
  - also referred to as unobserved effects or unobserved heterogeneity
  - ε_{it} = idiosyncratic error
- FEs subsume all time invariant x's
- FEs capture all time invariant attributes – observable and unobservable – of individual i

- pooled OLS equivalent to estimating

$$\begin{aligned} y_{it} &= \alpha + x_{it}\beta + \widetilde{\varepsilon}_{it} \\ \widetilde{\varepsilon}_{it} &= (\alpha_i - \alpha) + \varepsilon_{it} \end{aligned}$$

  where $\widetilde{\varepsilon}_{it}$ is known as a *composite error*

  – unbiasedness of $\widehat{\beta}_{POLS}$ requires $\text{Cov}(\alpha_i, x_{it}) = 0$ and $\text{Cov}(\varepsilon_i, x_{it}) = 0$

  – bias due to $\text{Cov}(\alpha_i, x_{it}) \neq 0$ known as *heterogeneity bias*

- pulling out α_i from the error term permits unbiased estimates of β even if $\text{Cov}(\alpha_i, x_{it}) \neq 0$

- if $\text{Cov}(\alpha_i, x_{it}) = 0$, then *random effects estimation* is more efficient

- if $\alpha_i = \alpha\ \forall i$, then pooled OLS is more efficient

- estimation methods when $\text{Cov}(\alpha_i, x_{it}) \neq 0$

  – LSDV (Least Squares Dummy Variable Model)

  – FD (first-differencing)

  – mean-differencing (FE estimator; within estimator)

- STATA: -*xtreg, fe fd*-, -*areg*-

- LSDV (Least Squares Dummy Variable Model)

$$y_{it} = \sum_{j=1}^{N} \alpha_j D_{ji} + x_{it}\beta + \varepsilon_{it}$$

where

$$D_{ji} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

- amounts to including N dummy vars, 1 for each group
- estimated by pooled OLS
- $\widehat{\beta}_{LSDV}$ is consistent even if $\text{Cov}(\alpha_i, x_{it}) \neq 0$ (regressors can always be correlated)
- only feasible computationally if N is of reasonable size

- FD (first-differencing)

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

  - implies

$$\begin{aligned} y_{i1} &= \alpha_i + x_{i1}\beta + \varepsilon_{i1} \\ &\vdots \\ y_{iT} &= \alpha_i + x_{iT}\beta + \varepsilon_{iT} \end{aligned}$$

  - taking differences between consecutive years yields

$$\begin{aligned} y_{i2} - y_{i1} &= (x_{i2} - x_{i1})\beta + (\varepsilon_{i2} - \varepsilon_{i1}) \\ &\vdots \\ y_{iT} - y_{iT-1} &= (x_{iT} - x_{iT-1})\beta + (\varepsilon_{iT} - \varepsilon_{iT-1}) \end{aligned}$$

    or, using new notation,

$$\begin{aligned} \Delta y_{i2} &= \Delta x_{i2}\beta + \Delta\varepsilon_{i2} \\ &\vdots \\ \Delta y_{iT} &= \Delta x_{iT}\beta + \Delta\varepsilon_{iT} \end{aligned}$$

    where Δ represents the change from the preceding year

  - the model to be estimated is

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta\varepsilon_{it},\ i = 1, ..., N; t = 2, ..., T$$

- notes
  - FD the data, then regress Δy_{it} on Δx_{it} using $N(T-1)$ observations
  - interpretation of $\widehat{\beta}_{FD}$ is same as original β
  - differencing eliminates α_i and any time invariant x's
  - consistency requires $\text{Cov}(\Delta x_{it}, \Delta\varepsilon_{it}) = 0$
    - known as *strict exogeneity*
    - requires $\text{Cov}(x_{it}, \varepsilon_{it}) = \text{Cov}(x_{it}, \varepsilon_{it-1}) = \text{Cov}(x_{it}, \varepsilon_{it+1}) = 0 \ \forall t$
  - estimator of FEs

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

  which is unbiased, but consistency requires $T \longrightarrow \infty$

- mean-differencing (FE estimator; within estimator)

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

  - implies

$$\overline{y}_i = \alpha_i + \overline{x}_i\beta + \overline{\varepsilon}_i$$

    and

$$\overline{\overline{y}} = \overline{\overline{\alpha}} + \overline{\overline{x}}\beta + \overline{\overline{\varepsilon}}$$

    where bars indicate average over T obs within group; double bars indicate average over entire sample

  - taking differences yields

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

    or, using new notation,

$$\widetilde{y}_{it} = \widetilde{x}_{it}\beta + \widetilde{\varepsilon}_{it}$$

– alternative representation

$$\widetilde{y}_{it} + \overline{\overline{y}} = \overline{\overline{\alpha}} + (\widetilde{x}_{it} + \overline{\overline{x}})\beta + (\widetilde{\varepsilon}_{it} + \overline{\overline{\varepsilon}})$$

or, using new notation,

$$\widetilde{\widetilde{y}}_{it} = \widetilde{\widetilde{x}}_{it}\beta + \widetilde{\widetilde{\varepsilon}}_{it}$$

– notes

∗ demean the data, then regress $\widetilde{y}_{it}$ on $\widetilde{x}_{it}$, or $\widetilde{\widetilde{y}}$ on $\widetilde{\widetilde{x}}$, using NT observations

∗ need to adjust degrees of freedom due to estimation of means

∗ interpretation of $\widehat{\beta}_{FE}$ is same as original β

∗ differencing eliminates α_i and any time invariant x's

∗ consistency requires $\text{Cov}(\widetilde{x}_{it}, \widetilde{\varepsilon}_{it}) = 0$

· again, *strict exogeneity*

· requires x_{it} to be independent of error term from *every* time period

- comparisons
  - LSDV and mean-differencing are identical
  - $T = 2 \Longrightarrow$ all three are identical
  - $T > 3 \Longrightarrow$ different, but both unbiased
- extensions
  - time dummies (DID estimator)

5.3 Random Effects

- motivation
  - if $\text{Cov}(x_{it}, \alpha_i) = 0$, then estimating N parameters α_i is inefficient (equivalently, loosing N obs to FD or MD is inefficient)
  - but, if $\alpha_i \neq \alpha\ \forall i$, then pooled OLS yields incorrect std errors since $\text{Cov}(\widetilde{\varepsilon}_{it}, \widetilde{\varepsilon}_{it'}) \neq 0$ (within-group serial correlation)

$$\text{Cov}(\widetilde{\varepsilon}_{it}, \widetilde{\varepsilon}_{it'}) = \begin{cases} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \text{if } t = t' \\ \sigma_\alpha^2 & \text{if } t \neq t' \end{cases}$$

    which implies *positive* serial correlation within groups
  - solution
    * leave α_i as part of the composite error
    * transform the data to a model with serially uncorrelated errors
    * known as Generalized Least Squares (GLS) estimation in general, RE estimation in this special case

- same setup as before

$$y_{it} = \alpha + x_{it}\beta + \widetilde{\varepsilon}_{it}, \quad \widetilde{\varepsilon}_{it} \sim N(0, \sigma_\varepsilon^2)$$

  where $\widetilde{\varepsilon}_{it}$ is the composite error term

- assume $\alpha_i \overset{iid}{\sim} N(0, \sigma_\alpha^2) =$ RE for group i

- assume $\text{Cov}(\alpha_i, \widetilde{\varepsilon}_{it}) = 0\ \forall t$

- RE estimation
  - transform the data to a model with serially uncorrelated errors
  - RE covariance structure

$$\underbrace{\Sigma_i}_{TxT} = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & & & \\ \sigma_\alpha^2 & \ddots & & \\ \vdots & \ddots & \ddots & \\ \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

and

$$\underbrace{\Sigma}_{NTxNT} = \begin{bmatrix} \Sigma_1 & & & \\ 0 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \Sigma_N \end{bmatrix}$$

  - define

$$\lambda = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_\alpha^2 + \sigma_\varepsilon^2}} \in [0,1]$$

  - λ-difference the data

$$y_{it} - \lambda \overline{y}_i = (x_{it} - \lambda \overline{x}_i)\beta + (\varepsilon_{it} - \lambda \overline{\varepsilon}_i)$$

implies $\mu_{it} \equiv \varepsilon_{it} - \lambda \overline{\varepsilon}_i \overset{iid}{\sim} N(0, \sigma_\mu^2)$

- steps
  - estimate the model using fixed effect methods or pooled OLS
  - obtain an estimate, $\widehat{\lambda}$
  - difference the data using $\widehat{\lambda}$
  - regress $y_{it} - \widehat{\lambda}\overline{y}_i$ on $x_{it} - \widehat{\lambda}\overline{x}_i$
- notes
  - special cases
    - $\lambda = 0 \Longrightarrow$ pooled OLS
    - $\lambda = 1 \Longrightarrow$ FE estimation
  - RE allows time invariant x's
  - consistency requires $\text{Cov}(\alpha_i, x_{it}) = \text{Cov}(\varepsilon_{it}, x_{it}) = 0$
- STATA: *-xtreg, re-*

5.4 Specification Tests

- Hausman test of FE vs. RE
  - intuition
    - if $\text{Cov}(\alpha_i, x_{it}) = 0$, then RE and FE are both consistent, but RE is more efficient

      $\Longrightarrow \widehat{\beta}_{RE} \approx \widehat{\beta}_{FE}$
    - if $\text{Cov}(\alpha_i, x_{it}) \neq 0$, then RE is inconsistent, but FE is consistent

      $\Longrightarrow \widehat{\beta}_{RE} \neq \widehat{\beta}_{FE}$
  - define test statistic based on difference $\widehat{\beta}_{FE} - \widehat{\beta}_{RE}$

    $$H = T\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right)'\left(\widehat{\Sigma}_{FE} - \widehat{\Sigma}_{RE}\right)^{-1}\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right) \sim \chi^2_K$$

    where $K = \#$ of x's
  - if test statistic is too large, then reject $\text{Cov}(\alpha_i, x_{it}) = 0$
  - STATA: -*hausman*-

- RE vs. pooled OLS

  – hypothesis

$$H_o \ : \ \sigma_\alpha^2 = 0$$
$$H_1 \ : \ \sigma_\alpha^2 \neq 0$$

  – Breusch-Pagan (1980) test

$$\lambda_{LM} = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^{N} \left(T \overline{\widehat{\varepsilon}}_i \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2} - 1 \right]^2 \sim \chi_1^2$$

  – STATA: -*xttest1*- after -*xtreg, re*-

- groupwise heteroskedasticity
  - errors are homoskedastic within groups, heteroskedastic across groups
  - e.g., errors for a given individual have same variance in each period, but each individual has a unique variance
  - structure:

$$\underbrace{\Sigma_i}_{TxT} = \begin{bmatrix} \sigma_i^2 & & & \\ 0 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \sigma_i^2 \end{bmatrix}$$

and

$$\underbrace{\Sigma}_{NTxNT} = \begin{bmatrix} \Sigma_1 & & & \\ 0 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \Sigma_N \end{bmatrix}$$

  - hypothesis

$$H_o : \sigma_i^2 = \sigma^2 \ \forall i$$

$$H_1 : \sigma_i^2 \neq \sigma^2 \text{ for some } i$$

  - modified Wald test statistic

$$W' = \sum_{i=1}^{N} \frac{\left(\widehat{\sigma}_i^2 - \widehat{\sigma}^2\right)^2}{V_i} \sim \chi_N^2$$

where

$$
\begin{aligned}
V_i &= \frac{1}{T-1}\sum\nolimits_{t=1}^{T}\left(\widehat{\varepsilon}_{it}^2 - \widehat{\sigma}_i^2\right)^2 \\
\widehat{\sigma}_i^2 &= \frac{1}{T}\sum\nolimits_{t=1}^{T}\widehat{\varepsilon}_{it}^2
\end{aligned}
$$

– notes

  * valid in presence of non-normality

  * lower power in 'large N, small T' FE models

– STATA: -*xttest3*- after -*xtreg, fe*-

- cross-sectional dependence

$$H_o \ : \ \text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0 \ \forall i \neq j$$
$$H_1 \ : \ \text{Cov}(\varepsilon_{it}, \varepsilon_{jt}) \neq 0 \text{ for some } i \neq j$$

  - $T > N$

    * Breusch-Pagan (1980) test

$$\lambda_{LM} = T \sum_{i=2}^{N} \sum_{j=1}^{i=1} \widehat{\rho}_{ij}^2 \sim \chi_d^2$$

      where

      · $d = N(N-1)/2$

      · $\widehat{\rho}_{ij} = \text{Corr}(\widehat{\varepsilon}_i, \widehat{\varepsilon}_j)$, $i \neq j$; specifically,

$$\widehat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt}}{\sqrt{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2\right)}\sqrt{\left(\sum_{t=1}^{T} \widehat{\varepsilon}_{jt}^2\right)}}$$

    * intuition

      · compute NxN correlation matrix

$$R = \begin{bmatrix} \rho_{11} & \cdots & \cdots & \rho_{1N} \\ \vdots & \ddots & & \\ \vdots & & \ddots & \\ \rho_{1N} & \cdots & \cdots & \rho_{NN} \end{bmatrix}$$

      · no correlation $\Longrightarrow R = I_N$

* test does not have good statistical properties when $T < N$, and likely to do worse as $N \longrightarrow \infty$
* STATA: -*xttest2*- after -*xtreg, fe*-

- $T < N$
  * Peasaran (2004) test

$$\lambda_{CD} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{i=N} \widehat{\rho}_{ij} \sim N(0,1)$$

  * STATA: -*xtcsd, pes*- after -*xtreg, fe re*-

5.5 Dynamic Panel Model

- model

$$y_{it} = x_{it}\beta + \gamma y_{it-1} + \alpha_i + \varepsilon_{it}, \quad \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

  where $i = 1, ..., N$; $t = 2, ..., T$; and $T \geqslant 3$

- estimation

  - even if $\text{Cov}(\alpha_i, x_{it}) = 0$, RE not applicable since $\text{Cov}(\alpha_i, y_{it-1}) \neq 0$
  - need FE/FD estimator
  - FD $\Longrightarrow$

$$\Delta y_{it} = \Delta x_{it}\beta + \gamma \Delta y_{it-1} + \Delta\varepsilon_{it}, \ i = 1, ..., N; t = 3, ..., T$$

  - but this model is not estimable by OLS since $\text{Cov}(\Delta\varepsilon_{it}, \Delta y_{it-1}) \neq 0$ since $\text{Cov}(\varepsilon_{it-1}, y_{it-1}) \neq 0$

- solutions
  - FD, then estimate via IV, treating Δy_{it-1} as endogenous
  - what are potential IVs?
    * need vars that are correlated with Δy_{it-1}, uncorrelated with $\Delta \varepsilon_{it}$
    * suitable candidates
      · x_{it-2} (through y_{it-2})
      · y_{it-2} (through y_{it-2})
      · $y_{it-3}, y_{it-4}, y_{it-5}, \ldots$ (through autoregressive process) ... e.g.,

$$\begin{aligned} \text{Cov}(\Delta y_{it-1}, y_{it-3}) &= \text{Cov}(y_{it-1}, y_{it-3}) - \text{Cov}(y_{it-2}, y_{it-3}) \\ &\neq 0 \end{aligned}$$

    * lots of instruments (beware of weak IVs)
  - simple solution: FD, then use TSLS with x_{it-2}, y_{it-2} as IVs

*

– more complex solution

* estimation by GMM to utilize more instruments

  · writing out model for each period yields

$$\begin{aligned}\Delta y_{i3} &= \Delta x_{i3}\beta + \gamma\Delta y_{i2} + \Delta\varepsilon_{i3} \\ &\vdots \\ \Delta y_{iT} &= \Delta x_{iT}\beta + \gamma\Delta y_{iT-1} + \Delta\varepsilon_{iT}\end{aligned}$$

  where IVs for Δy_{i2} are x_{i1}, y_{i1}; IVs for Δy_{i3} are x_{i2}, y_{i2}, y_{i1}; ... ; IVs for Δy_{iT-1} are $x_{iT-2}, y_{iT-2}, ..., y_{i2}, y_{i1}$

  · not usual TSLS set-up

  · GMM allows moment conditions to be derived using as many IVs as desired

  · requires ε_{it} to be serially uncorrelated; or, equivalently, $\Delta\varepsilon_{it}$ should be $AR(1)$

* STATA: -*xtabond*- (Arellano & Bond 1991)

- persistence
  - Blundell & Bond (1998) show that if $|\gamma| > 0.8$ or so, TSLS and A-B estimator do not work very well (weak IVs)
  - solution
    * add additional moment conditions derived from the model in levels

$$y_{it} = x_{it}\beta + \gamma y_{it-1} + \alpha_i + \varepsilon_{it}$$

    * what are IVs for y_{it-1}?
      · Δy_{it-1} (independent of α_i, ε_{it})
      · Δx_{it-1} (independent of α_i, ε_{it})
      · $\Delta y_{it-2}, \Delta y_{it-3}, ...$ (through autoregressive process)
  - STATA: -*xtabond2*- (system estimator)