EC968

Panel Data Analysis

Steve Pudney

ISER

Lecture 3: Endogeneity and Instrumental Variables

Static models: types of endogeneity
- Within- and between-group IV estimators
- The Hausman-Taylor approach

Dynamic regression
- IV and GMM estimators

Endogeneity in static models

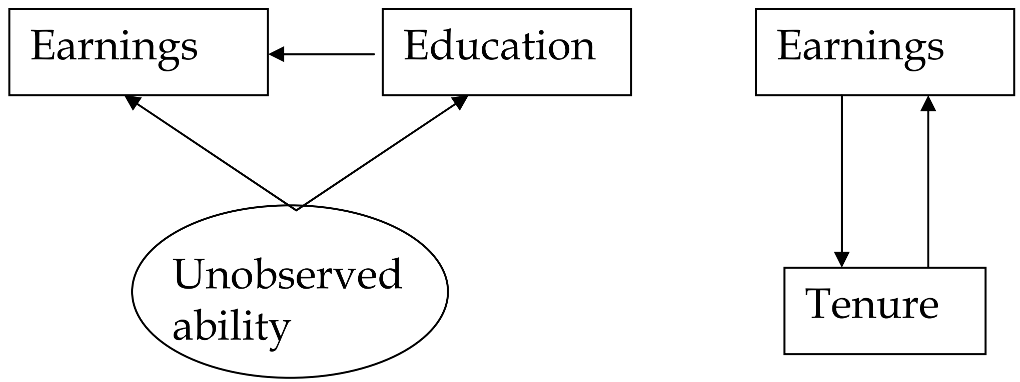
Example: an earnings model

$$y_{it} = \alpha_1 \, Educ_i + \alpha_2 \, Female + \beta_1 \, Age_{it} + \beta_2 \, Tenure_{it} + u_i + \varepsilon_{it}$$

Two forms of endogeneity:

Two-way causation: experience is rewarded with high pay & workers tend to stay in high-paid jobs

Unobserved common factors: ability is rewarded with high pay & high-ability people stay longer in education



(a) unobserved common factor

(b) 2-way causation

Example of endogeneity

Example: an earnings model

$$y_{it} = \alpha_1\, Educ_i + \alpha_2\, Female + \beta_1\, Age_{it} + \beta_2\, Tenure_{it} + u_i + \varepsilon_{it}$$

(1) Two-way causation: workers tend to stay in high-paid jobs:

Tenure model: $Tenure_{it} = \gamma y_{it} + \upsilon_{it} \qquad (\gamma > 0)$

$$= \gamma(\alpha_1\, Educ_i + \ldots + \beta_1\, Age_{it} + \beta_2\, Tenure_{it} + u_i + \varepsilon_{it}) + \upsilon_{it}$$

$$= [\, \gamma(\alpha_1\, Educ_i + \ldots + \beta_1\, Age_{it} + u_i + \varepsilon_{it}) + \upsilon_{it}\,] \,/\, (1 - \gamma\beta_2)$$

$$\Rightarrow \qquad \mathrm{cov}(Tenure_{it}, u_i) = \gamma\sigma_u^2 / (1 - \gamma\beta_2)$$

$$\mathrm{cov}(Tenure_{it}, \varepsilon_{it}) = \gamma\sigma_\varepsilon^2 / (1 - \gamma\beta_2)$$

(2) Unobserved common factors: u_i represents ability & high-ability people stay longer in education:

$Educ_i = \delta u_i$ + other vars $\qquad (\delta > 0)$

$$\Rightarrow \qquad \mathrm{cov}(Educ_i, u_i) = \delta\sigma_u^2$$

$$\mathrm{cov}(Educ_i, \varepsilon_{it}) = 0$$

Strategy for dealing with endogeneity

Type of endogeneity	Consequences	Method
2-way causation (e.g. tenure → wage & wage → tenure)	$\text{Cov}(x,u) \neq 0$ $\text{Cov}(x,\varepsilon) \neq 0$	Within-group IV (w-g to eliminate u_i and IV to deal with covariance with ε)
Common unobserved factor which persists over time (e.g. ability → wage, ability → education & education → wage)	$\text{Cov}(x,u) \neq 0$ $\text{Cov}(x,\varepsilon) = 0$	Within-group regression (eliminates u_i) and Hausman-Taylor to estimate coefficients of $\mathbf{z}_i$
Common unobserved factor which does not persist over time (e.g. job loss → wage & job loss → tenure)	$\text{Cov}(x,u) = 0$ $\text{Cov}(x,\varepsilon) \neq 0$	Random-effects IV, using as IVs variables which are correlated with risk of job loss but not wages; no need to use within-group, since u_i isn't correlated with x
None	$\text{Cov}(x,u) = 0$ $\text{Cov}(x,\varepsilon) = 0$	GLS random effects regression

The Instrumental Variables principle

Simple example – a cross-section regression model:

$$y_i = x_i \beta + \varepsilon_i$$

Problem: simultaneous causation

$\Rightarrow \mathrm{cov}(x_i\,,\,\varepsilon_i) \neq 0$

$\Rightarrow$ OLS regression of y_i on x_i is biased

But assume there is another variable q_i with two properties:

Validity: $\mathrm{cov}(q_i\,,\,\varepsilon_i) = 0$

Relevance: $\mathrm{cov}(q_i\,,\,x_i) \neq 0$

The *validity* requirement says that the instrument must not suffer from the same endogeneity problem that x_i does;

The *relevance* requirement says that the instrument must be closely related to x_i

Motivation for the IV method

The assumption of instrument validity is a ***moment condition*** which states that a particular *moment*, cov(q, ε), must be equal to zero

But the model tells us that: $\varepsilon_i = y_i - x_i \beta$, so:

$$\begin{aligned} \text{cov}(q_i, \varepsilon_i) &= \text{cov}(q_i, [\, y_i - x_i \beta \,]) \\ &= \text{cov}(q_i, y_i) - \beta\, \text{cov}(q_i, x_i) \\ &= 0 \quad \text{(instrument validity requirement)} \end{aligned}$$

Solve for β:

$$\beta = \text{cov}(q_i, y_i) \;/\; \text{cov}(q_i, x_i)$$

So, if q is a valid instrument, β must be equal to the ratio of the population covariance between q and y and between q and x.

The simple Instrumental Variable (IV) estimator

The sample analogue of this moment condition provides an estimator:

$$\hat{\beta}_{IV} = \frac{\text{sample cov}(q, y)}{\text{sample cov}(q, x)} = \frac{\sum_{i=1}^{n}(q_i - \bar{q})(y_i - \bar{y})}{\sum_{i=1}^{n}(q_i - \bar{q})(x_i - \bar{x})}$$

This can be generalised to:

- More than one explanatory variable in ($\mathbf{z}_i$, $\mathbf{x}_{it}$)
- More than one instrumental variable:

$$\hat{\boldsymbol{\beta}}_{IV} = \left(\mathbf{X'Q}(\mathbf{Q'Q})^{-1}\mathbf{Q'X}\right)^{-1}\mathbf{X'Q}(\mathbf{Q'Q})^{-1}\mathbf{Q'y}$$

- Require no. instruments ≥ no. explanatory variables

Simultaneity: Within-group IV estimation

Model:

$$y_{it} = \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

Partition $\mathbf{x}_{it}$:

$$\mathbf{x}_{it} = (\mathbf{x}_{1it}, \mathbf{x}_{2it}),$$

where: $\text{cov}(\mathbf{x}_{1it}, \varepsilon_{it}) = 0$ and $\text{cov}(\mathbf{x}_{2it}, \varepsilon_{it}) \neq 0$

Instruments $\mathbf{q}_{2it}$ (at least as many as in $\mathbf{x}_{2it}$)

Full IV vector $\mathbf{q}_{it} = (\mathbf{x}_{1it}, \mathbf{q}_{2it})$

Within-group transformation:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i$$

IV estimator:

$$\hat{\boldsymbol{\beta}}_{WIV} = \left(\mathbf{W}_{xq}\mathbf{W}_{qq}^{-1}\mathbf{W}_{qx}\right)^{-1}\mathbf{W}_{xq}\mathbf{W}_{qq}^{-1}\mathbf{w}_{qy}$$

Consistency

$$\underset{n\to\infty}{\text{plim}}\,\hat{\boldsymbol\beta}_{WIV} = \boldsymbol\beta + \left(\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{W}_{xq}\left(\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{W}_{qq}\right)^{-1}\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{W}_{qx}\right)^{-1}\times$$

$$\left(\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{W}_{xq}\left(\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{W}_{qq}\right)^{-1}\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{w}_{q\varepsilon}\right)$$

$$= \boldsymbol\beta$$

This consistency property holds because:

- The within-group transform removes u_i, which may be correlated with $\mathbf{x}_{2it}$
- The instruments are uncorrelated with ε, so:

$$\underset{n\to\infty}{\text{plim}}\frac{1}{n}\mathbf{w}_{q\varepsilon} = \underset{n\to\infty}{\text{plim}}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(\mathbf{q}_{it} - \overline{\mathbf{q}}_i)'(\varepsilon_{it} - \overline{\varepsilon}_i) = \mathbf{0}$$

Between-group and random-effects IV estimators

Analogous to the regression case:

$$\hat{\boldsymbol{\beta}}_{BIV} = \left(\mathbf{B}_{x^*q}\mathbf{B}_{qq}^{-1}\mathbf{B}_{qx^*}\right)^{-1}\mathbf{B}_{x^*q}\mathbf{B}_{qq}^{-1}\mathbf{b}_{qy}$$

$$\hat{\boldsymbol{\beta}}_{REIV} = \left(\mathbf{R}_{x^*q}\mathbf{R}_{qq}^{-1}\mathbf{R}_{qx^*}\right)^{-1}\mathbf{R}_{x^*q}\mathbf{R}_{qq}^{-1}\mathbf{r}_{qy}$$

where $\mathbf{x}_{it}^* = \left(\mathbf{z}_i, \mathbf{x}_{it}\right)$,

$$\mathbf{R}_{x^*q} = \sum_{i=1}^{n}\sum_{t=1}^{T_i}\left(\mathbf{x}_{it}^* - \theta_i\overline{\mathbf{x}}_i^*\right)'\left(\mathbf{q}_{it} - \theta_i\overline{\mathbf{q}}_i\right), \quad \text{etc.}$$

and $\theta_i = 1 - \sqrt{\sigma_\varepsilon^2 / \left(\sigma_\varepsilon^2 + T_i\sigma_u^2\right)}$

If $\mathrm{cov}(\mathbf{q}_{it}, u_i) \neq 0$, then both $\hat{\boldsymbol{\beta}}_{BIV}$ and $\hat{\boldsymbol{\beta}}_{REIV}$ are inconsistent $\Rightarrow$ a stronger requirement for instrument validity

Simultaneity involving only individual effects: the Hausman-Taylor case

Model:

$$y_{it} = \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it}$$

Partition $\mathbf{x}_{it}$ and $\mathbf{z}_i$:

$$\mathbf{x}_{it} = (\mathbf{x}_{1it}, \mathbf{x}_{2it}), \quad \mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i}),$$

where:

$$\mathrm{E}(u_i \mid \mathbf{x}_{1it}) = 0, \ \mathrm{E}(u_i \mid \mathbf{z}_{1i}) = 0$$

$$\mathrm{E}(u_i \mid \mathbf{x}_{2it}) \neq 0, \ \mathrm{E}(u_i \mid \mathbf{z}_{2i}) \neq 0$$

But we must assume:

$\mathrm{E}(\varepsilon_{it} \mid \mathbf{x}_{it}) = 0, \ \mathrm{E}(\varepsilon_{it} \mid \mathbf{z}_i) = 0$ for all x- and z-variables

(no simultaneous determination of y_{it} and $(\mathbf{z}_i, \mathbf{x}_{it})$ **!!!!**)

Identification condition: dim $(\mathbf{x}_{1it}) \geq$ dim $(\mathbf{z}_{2i})$

Method: use $\mathbf{x}_{1it}$ as IVs for $\mathbf{z}_{2i}$

The Hausman-Taylor (1981) estimator

Step 1: compute the within-group estimator for $\boldsymbol{\beta}$:

$$\Rightarrow \text{ regress } \quad y_{it} - \bar{y}_i \text{ on } \quad \mathbf{x}_{it} - \bar{\mathbf{x}}_i \Rightarrow \quad \hat{\boldsymbol{\beta}}_W$$

Step 2: construct within-group residuals & estimate σ_ε^2 :

$$\hat{\varepsilon}_{it} = y_{it} - \bar{y}_i - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}}_W$$

$$\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^{n}\sum_{t=1}^{T_i} \hat{\varepsilon}_{it}^{\,2} \Big/ \left(n\left(\bar{T} - 1\right) - k_x\right)$$

Step 3: estimate model for $\quad \hat{e}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_W$:

$$\hat{e}_i = \alpha_0 + \mathbf{z}_i\boldsymbol{\alpha} + \text{residual}, \qquad i = 1 \ldots n, \; t = 1 \ldots T_i$$

use as IVs $\mathbf{q}_{it}$ = $[\mathbf{x}_{1it}, \mathbf{z}_{1i}]$ so: $\quad \hat{\boldsymbol{\alpha}} = \left(\mathbf{B}_{zq}\mathbf{B}_{qq}^{-1}\mathbf{B}_{qz}\right)^{-1}\mathbf{B}_{zq}\mathbf{B}_{qq}^{-1}\mathbf{b}_{q\hat{e}}$

Step 4: Construct $\quad \hat{e}_i^* = \bar{y}_i - \mathbf{z}_i\hat{\boldsymbol{\alpha}} - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_W$; estimate σ_u^2 from $\hat{\varepsilon}_{it}$ and $\hat{e}_i^*$

Step 5: Carry out the random effects transform and estimate:

$$\left(y_{it} - \theta_i\bar{y}_i\right) = \mathbf{z}_i(1 - \theta_i)\boldsymbol{\alpha} + \left(\mathbf{x}_{it} - \theta_i\bar{\mathbf{x}}_i\right)\boldsymbol{\beta} + \left(\varepsilon_{it} - \theta_i\bar{\varepsilon}_i\right)$$

using as IVs $\quad \mathbf{q}_{it} = \left[\mathbf{z}_{1i}, \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right), \bar{\mathbf{x}}_{1i}\right]$

(NB more elaborate IVs can be used, see Amemiya-MacCurdy, 1986).

Endogeneity: BHPS examples

Model:

$$Ln\ wage = \alpha_0 + \alpha_1\ Female + \alpha_2\ Education\ beyond\ GCSE$$
$$+ \beta_1\ Age + \beta_2\ Job\ tenure + u + \varepsilon$$

(1) Is job tenure jointly determined with the wage?

- Use the standard IV/2SLS estimator in w-g, b-g or r-e form
- Possible instruments: *Married, Spouse part-time, Spouse full-time, Dissatisfied with hours,*
- But are these valid instruments?

(2) Is educational attainment influenced by the same unobservable factors as labour market success?

- Use the Hausman-Taylor estimator
- Instruments come from within the model
- But is everything uncorrelated with ε?

Within-group regression

```
. xtreg logearn age postGCSE tenure, fe

Fixed-effects (within) regression               Number of obs      =     38404
Group variable (i): pid                         Number of groups   =      7700

R-sq:  within  = 0.0983                         Obs per group: min =         1
       between = 0.0024                                        avg =       5.0
       overall = 0.0038                                        max =        11

                                                F(3,30701)         =   1115.13
corr(u_i, Xb)  = -0.4195                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     logearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0249189   .0004778    52.16   0.000     .0239824    .0258554
    postGCSE |   .0263467   .0089311     2.95   0.003     .0088413     .043852
      tenure |   .0016804   .0004299     3.91   0.000     .0008377     .002523
       _cons |   .9805382   .0174738    56.11   0.000     .9462889    1.014787
-------------+----------------------------------------------------------------
     sigma_u |  .54846498
     sigma_e |  .24922759
         rho |  .82885214   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(7699, 30701) =    14.66         Prob > F = 0.0000
```

Within-group IV estimates

```
. xtivreg logearn age postGCSE (tenure = dumm*), fe
note: dumm6 dropped due to collinearity
Fixed-effects (within) IV regression            Number of obs      =     38404
Group variable: pid                             Number of groups   =      7700

R-sq:  within  = 0.0974                         Obs per group: min =         1
       between = 0.0027                                        avg =       5.0
       overall = 0.0040                                        max =        11

                                                Wald chi2(3)       =  2.40e+06
corr(u_i, Xb)  = -0.4164                        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     logearn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      tenure |   .0039841    .007105     0.56   0.575    -.0099415    .0179097
         age |   .0243511   .0018121    13.44   0.000     .0207995    .0279027
    postGCSE |   .0279968   .0102783     2.72   0.006     .0078518    .0481418
       _cons |   .9909042   .0363862    27.23   0.000     .9195886     1.06222
-------------+----------------------------------------------------------------
     sigma_u |  .54731645
     sigma_e |  .24934411
         rho |  .82812356   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F  test that all u_i=0:     F(7699,30701) =    14.63      Prob > F    = 0.0000
------------------------------------------------------------------------------
Instrumented:   tenure
Instruments:    age postGCSE dumm1-dumm12
```

Hausman test comparing w-g OLS & IV

```
. hausman olsfe ivfe

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     olsfe         ivfe         Difference          S.E.
-------------+----------------------------------------------------------------
         age |    .0249189     .0243511        .0005678               .
    postGCSE |    .0263467     .0279968       -.0016501               .
      tenure |    .0016804     .0039841       -.0023038               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
          B = inconsistent under Ha, efficient under Ho; obtained from xtivreg

    Test:  Ho:  difference in coefficients not systematic

                 chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =        0.11
               Prob>chi2 =      0.9912
```

Endogeneity of education: Hausman-Taylor

```
. xthtaylor logearn age tenure postGCSE2 female cohort, endog(tenure postGCSE2)
Hausman-Taylor estimation                       Number of obs      =     38404
Group variable (i): pid                         Number of groups   =      7700
                                                Obs per group: min =         1
                                                               avg =       5.0
                                                               max =        11
Random effects u_i ~ i.i.d.                     Wald chi2(5)       =   4111.99
                                                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
     logearn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         age |   .0253258   .0004155    60.95   0.000     .0245115    .0261402
TVendogenous |
      tenure |   .0016367   .0003903     4.19   0.000     .0008717    .0024016
TIexogenous  |
      female |  -.1749879   .0436307    -4.01   0.000    -.2605026   -.0894732
      cohort |   .0115968   .0033232     3.49   0.000     .0050834    .0181102
TIendogenous |
   postGCSE2 |   1.260647   .3184888     3.96   0.000     .6364202    1.884873
             |
       _cons |  -22.45571   6.338539    -3.54   0.000    -34.87902   -10.03241
-------------+----------------------------------------------------------------
     sigma_u |  1.7227596
     sigma_e |  .24925073
         rho |  .97949657   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Dynamic models for continuous dependent variables

Adjustment may be imperfect – how to model it? Any conventional time-series model can be used, *e.g.* AR(1):

$$y_{it} = \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + \gamma y_{it\text{-}1} + u_i + \varepsilon_{it} \qquad (1)$$

or static model with AR(1) errors:

$$y_{it} = \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + u_i + \varepsilon_{it} \qquad (2)$$

$$\varepsilon_{it} = \rho\, \varepsilon_{it\text{-}1} + \eta_{it}$$

$$\Rightarrow \quad y_{it} = \mathbf{z}_i (1\text{-}\rho)\boldsymbol{\alpha} + (\mathbf{x}_{it} \text{-} \rho\, \mathbf{x}_{it\text{-}1})\boldsymbol{\beta} + \rho\, y_{it\text{-}1} + (1\text{-}\rho)u_i + \eta_{it} \qquad (2')$$

NB: model (1) implies gradual adjustment to change in **x**; model (2) implies a full immediate response.

More general distributed lag models can be used (e.g. ECMs, ARMA, etc.)

Within-group estimation

Within-group transformed model (e.g. AR(1)):

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \gamma(y_{it-1} - \bar{y}_i^*) + \varepsilon_{it} - \bar{\varepsilon}_i$$

where:

$$\bar{y}_i^* = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it-1} = \frac{1}{T_i}\sum_{t=0}^{T_i-1} y_{it} \neq \bar{y}_i$$

NB we assume a compact panel (why?) and an observable initial condition y_{i0}

What are the statistical properties of a regression of $y_{it} - \bar{y}_i$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ and $(y_{it-1} - \bar{y}_i^*)$?

Properties of the within-group estimator

Consider the solved distributed lag form of (1):

$$y_{it} = \sum_{s=0}^{t-1} \gamma^s \left(\mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it-s} \boldsymbol{\beta} + u_i + \varepsilon_{t-s} \right) + \gamma^t y_{i0}$$

$$= \frac{1-\gamma^t}{1-\gamma} \left(\mathbf{z}_i \boldsymbol{\alpha} + u_i \right) + \sum_s \gamma^s \mathbf{x}_{it-s} \boldsymbol{\beta} + \left[\varepsilon_{it} + \gamma \varepsilon_{it-1} + \ldots + \gamma^{t-1} \varepsilon_{i1} \right] + \gamma^t y_{i0}$$

$\Rightarrow$ y_{it} is a function of $\varepsilon_{it} \ldots \varepsilon_{i1}$

$\Rightarrow$ $\bar{y}_i^* = \sum_{t=0}^{T_i-1} y_{it} \Big/ T_i$ is a function of $\varepsilon_{iT\text{-}1} \ldots \varepsilon_{i1}$ and y_{i0}

$\Rightarrow$ $y_{it} - \bar{y}_i^*$ is correlated with $\varepsilon_{it} - \bar{\varepsilon}_i$

$\Rightarrow$ bias in within-group regression coefficients

Bias-correction approaches

- Bias in the dynamic within-group regression estimator:
  - Complicated mathematical form (Nickell, 1981)
  - generally negative for γ for small T (even if true γ is zero)
  - Pooled OLS, b-g & random effects also biased.
- It is possible to construct an (approximately) bias-corrected within-group estimator, suitable even when n is only moderately large:
  - Bun & Kiviet (Ecs. Letters, 2003); Bruno (Ecs. Letters, 2005)
  - Stata module xtlsdvc (http://ideas.repec.org/c/boc/bocode/s450101.html)

A simple IV estimator

The within-group transform complicates estimation with lagged endogenous variables. Consider time-differencing:

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \gamma \Delta y_{it-1} + \Delta \varepsilon_{it} \ , \qquad t = 2 \ldots T_i \qquad (1)$$

The problem now is that the error term, $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it\text{-}1}$ is a MA(1) process which contains $\varepsilon_{it\text{-}1}$, which is correlated with $\Delta y_{it\text{-}1}$.

$\Rightarrow$ Find a set of instruments correlated with $\Delta y_{it\text{-}1}$ but uncorrelated with $\varepsilon_{it\text{-}1}$

$\Rightarrow$ All lagged $\mathbf{x}_{it}$ and $y_{it\text{-}2} \ldots y_{i0}$ are valid instruments if $\{\varepsilon_{it}\}$ is serially independent

$\Rightarrow$ Simplest IV estimator (Anderson Hsiao) estimates (1), using instruments ($\mathbf{x}_{it}$, $\mathbf{x}_{it\text{-}1}$, $\mathbf{x}_{it\text{-}2}$,$y_{it\text{-}2}$).

$\Rightarrow$ We can only use observations $t = 2 \ldots T_i$. Each extra lag used as an instrument loses us n observations.

$\Rightarrow$ Once $\hat{\boldsymbol{\beta}}_{IV}$ is found, estimate $\boldsymbol{\alpha}$ by regressing $\bar{y}_i - \bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}}_{IV} - \hat{\gamma}\,\bar{y}_i^*$ on $\mathbf{z}_i$

Problems with IV estimators

Suppose y_{it} is a random walk (e.g. Hall's (1978) form of the permanent income hypothesis: dynamic choice models based on Euler conditions).

$\Rightarrow y_{it\text{-}2}$ is uncorrelated with $\Delta y_{it\text{-}1}$ and is not a valid instrument

$\Rightarrow$IV methods based on a differenced model won't work well if there is a near-unit root

Any method based solely on the differenced equation ignores potentially valuable information contained in the initial condition y_{i0}

What is the optimal point on the trade-off between the number of lags used as instruments and the number of time periods retained in the estimation sample?

System estimators

The time-differenced model:

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \gamma \Delta y_{it-1} + \Delta \varepsilon_{it} \ , \qquad t = 2 \ldots T_i \qquad (1)$$

This is a system of T_i-1 linear equations with cross-correlated errors (since $\Delta\varepsilon_{it}$ is correlated with $\Delta\varepsilon_{it-1}$ and $\Delta\varepsilon_{it+1}$)

There is also some (related) process generating the initial conditions, y_{i0} and y_{i1}, which could provide further equations.

A different number of instruments is available for each of the equations in (1):

E.g. the equation for t = 2 has only ($\mathbf{x}_{i0} \ldots \mathbf{x}_{iT}$, y_{i0});

the equation for $t = T_i$ has ($\mathbf{x}_{i0} \ldots \mathbf{x}_{iT}$, $y_{i0} \ldots y_{iT-2}$).

NB it's assumed here that $\mathbf{x}_{i0}$ is observable

Generalised method of moments

IV estimators are members of the class of GMM estimators *e.g.* the 2SLS estimator, $\hat{\boldsymbol{\beta}}_{IV} = \left(\mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{y}$ is the following M-estimator:

$$\hat{\boldsymbol{\beta}}_{IV} = \arg\min_{\boldsymbol{\beta}} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)$$

$$= \arg\min_{\boldsymbol{\beta}} \hat{\mathbf{m}}(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta})'\mathbf{V}^{-1}\hat{\mathbf{m}}(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta})$$

where $\hat{\mathbf{m}}$ is the "sample analogue", $n^{-1}\mathbf{Q}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$, of a moment, $E\mathbf{q}'\varepsilon$, assumed to be zero in the population.

V is a weighting matrix proportional to the asymptotic covariance matrix of the moment condition (in this standard 2SLS example $\sigma_\varepsilon^{\,2}\mathbf{Q}'\mathbf{Q}$, where $\sigma_\varepsilon^{\,2}$ is the residual variance).

GMM can be extended to any number of moment conditions

Arellano-Bond GMM (1991)

We have T_i -2 differenced equations (1).

The instruments for equation t are:

$$\mathbf{q}_{it} = (\mathbf{x}_{i0} \ldots \mathbf{x}_{iT}, y_{i0} \ldots y_{it-2})$$

Full set of moment conditions:

$E\, \mathbf{q}_{i2}' \Delta\varepsilon_{i2} = 0$ $(T_i+1)k_x+1$ conditions

$E\, \mathbf{q}_{i3}' \Delta\varepsilon_{i3} = 0$ $(T_i+1)k_x+2$ conditions

.

.

$E\, \mathbf{q}_{iT}' \Delta\varepsilon_{iT} = 0$ $(T_i+1)k_x+T_i-1$ conditions

$\hat{\mathbf{m}}$ is a $[(T_i+1)(T_i-1)k_x + T_i(T_i-1)/2] \times 1$ moment vector

The optimal choice for $\mathbf{V}$ is $E\hat{\mathbf{m}}_i\hat{\mathbf{m}}_i'$

More conditions can be added (*e.g.* for $\mathbf{z}_i$ and to impose the homoskedasticity assumption on ε_{it}). ***But*** GMM often works badly in finite samples with many moment conditions.

Further developments: initial conditions

Arellano-Bond ignores the initial conditions y_{i0} and y_{i1} and only uses moment conditions for $\Delta y_{i2} \ldots \Delta y_{iT}$.

To progress further, we need additional assumptions about the initial conditions. One possibility is:

Equilibrium initial values. If the process is homogeneous and long-established:

$$y_{i0} = \frac{\mathbf{z}_i \boldsymbol{\alpha} + u_i}{1-\gamma} + \sum_{s=0}^{\infty} \gamma^s \left(\mathbf{x}_{i,-s} \boldsymbol{\beta} + \varepsilon_{i,-s} \right)$$

⇒Coefficient of u_i in equation for y_{i0} is $(1-\gamma)^{-1}$

⇒But the quantity $\sum_{s=0}^{\infty} \gamma^s \mathbf{x}_{i,-s}$ is unobserved

⇒Also, do people really have infinite pasts?

If lagged levels of y_{it} are poor instruments for $\Delta y_{it\text{-}1}$, can we go back to using the equations in level form?

Extended system methods

Arellano & Bover (1995) and Blundell & Bond (1998) (see also Bhargava & Sargan, 1983) suggested using the model in *both* differenced and levels form to generate GMM moment conditions.

Question: in the levels model

$$y_{it} = \mathbf{z}_i \boldsymbol{\alpha} + \mathbf{x}_{it} \boldsymbol{\beta} + \gamma y_{it\text{-}1} + u_i + \varepsilon_{it} \ ,$$

is there a good instrument for $y_{it\text{-}1}$? This instrument must be uncorrelated with u_i as well as ε_{it} .

A&B suggested $\Delta y_{it\text{-}1}$, *etc*.. The instrument validity condition is $E[\Delta y_{it\text{-}1} (u_i + \varepsilon_{it})] = 0$, which in turn requires (see B&B, 1998):

$$E\, u_i \,[\, y_{i0} - u_i/(1\text{-}\gamma)] = 0 \quad \text{and} \quad E\, u_i \Delta\varepsilon_{it} = 0$$

The former is a strong assumption but, if true, improves estimation precision dramatically in highly-persistent models (*i.e.* when $\gamma \approx 1$)

Example

Model:

$$Ln\ wage_{it} = \alpha_0 + \alpha_1\ Female_i + \alpha_2\ Education\ beyond\ GCSE_i + \alpha_2\ Cohort_i + \beta_1\ Age_{it} + \beta_2\ Job\ tenure_{it} + \gamma\ Ln\ wage_{it\text{-}1} + u_i + \varepsilon_{it}$$

Estimate by:

- Within-group regression
- Arellano-Bond
- Blundell-Bond

Note: this is a poor model

- Significant differences between Arellano-Bond & Blundell-Bond
- Significant Sargan χ^2 test for instrument validity for both
- Significant 2^{nd}-order autocorrelation in Blundell-Bond

$\Rightarrow$ Investigate higher-order dynamics, omitted variables, endogenity issues?

Within-group regression

```
xtreg logearn l.logearn age postGCSE2 tenure female cohort, fe

Fixed-effects (within) regression               Number of obs      =     25419
Group variable (i): pid                         Number of groups   =      5798

R-sq:  within  = 0.1407                         Obs per group: min =         1
       between = 0.1217                                        avg =       4.4
       overall = 0.1337                                        max =         9

                                                F(3,19618)         =   1070.88
corr(u_i, Xb)  = -0.0940                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     logearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logearn |
         L1. |   .1918289    .006843    28.03   0.000      .178416    .2052417
         age |   .0209275   .0005674    36.88   0.000     .0198154    .0220397
   postGCSE2 |  (dropped)
      tenure |   .0001329   .0004798     0.28   0.782    -.0008076    .0010734
      female |  (dropped)
      cohort |  (dropped)
       _cons |    .797273   .0220869    36.10   0.000     .7539808    .8405652
-------------+----------------------------------------------------------------
     sigma_u |  .45773502
     sigma_e |  .22598029
         rho |  .80403163   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Arellano-Bond

```
. xtabond2 logearn l.logearn age tenure, gmm(l.logearn) iv(age tenure)
noleveleq Favoring speed over space. See help matafavor.

Arellano-Bond dynamic panel-data estimation, one-step difference GMM results
------------------------------------------------------------------------------
Group variable: pid                             Number of obs      =     16769
Time variable : year                            Number of groups   =      4658
Number of instruments = 38                      Obs per group: min =         1
Wald chi2(3)   =    563.04                                     avg =      3.60
Prob > chi2    =     0.000                                     max =         7
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logearn |
         L1. |   .1265653   .0209072     6.05   0.000     .0855879    .1675427
         age |     .02477   .0014812    16.72   0.000     .0218669    .0276731
      tenure |  -.0001292   .0006587    -0.20   0.844    -.0014203    .0011618
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(35) = 155.31    Prob > chi2 =  0.000

Arellano-Bond test for AR(1) in first differences: z = -25.72  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   1.72  Pr > z =  0.085
------------------------------------------------------------------------------
```

Blundell-Bond

```
. xtabond2 logearn l.logearn age tenure, gmm(l.logearn) iv(age tenure)
Favoring speed over space. See help matafavor.

Arellano-Bond dynamic panel-data estimation, one-step system GMM results
------------------------------------------------------------------------------
Group variable: pid                             Number of obs      =     25419
Time variable : year                            Number of groups   =      5798
Number of instruments = 46                      Obs per group: min =         1
Wald chi2(3)  =    221.24                                      avg =      4.38
Prob > chi2   =     0.000                                      max =         9
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logearn |
         L1. |   .2873673   .0199367    14.41   0.000      .248292    .3264426
         age |   1.08e-06     .00028     0.00   0.997    -.0005477    .0005498
      tenure |  -.0017568   .0005668    -3.10   0.002    -.0028677   -.0006458
       _cons |   1.449055   .0395529    36.64   0.000     1.371533    1.526578
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(42) = 131.85    Prob > chi2 =  0.000

Arellano-Bond test for AR(1) in first differences: z = -38.07  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =   2.84  Pr > z =  0.005
------------------------------------------------------------------------------
```