Lecture 2

Panel Data Models (II)

(Revised 2010)

Read:
- Wooldridge ch 11
- Greene ch 13
- Verbeek ch 10

EC671

Lee

Outline

1. Endogeneity in panel data models

   - $Cov(X_{it}, e_{is}) = 0$,

     three cases ; $s=t$, $s=t+k$, $s=t-k$

   - Hausman-Taylor model

2. Dynamic Panel Data Models [Also, weakly exog.]

   - Three methods (FE, RE, FD) fail. why & when?

   - suggested solutions

     (FD-IV, GMM, System GMM..)

3. Other panel data Models

   - Extensions

   - Further considerations

Endogeneity in Panel Data Models

$y_{it} = x_{it}\beta + \boxed{w_{it}}\gamma + \alpha_i + \boxed{u_{it}}$

Case 1* $Cov(w_{it}, u_{it}) \neq 0$, but $cov(w_{it}, u_{is}) = 0$, $t \neq s$

.. contemporaneous correlation

; usual endogeneity issue (4 sources)

; solution = 2SLS (xtivreg or xtivreg2 in Stata)

Case 2** $Cov(w_{it}, u_{it}) = 0$, $cov(w_{it}, u_{is}) = 0$ for $s > t$ (uncorrelated with future errors)

; predetermined (feedback effect) $\Rightarrow$ But $Cov(w_{it}, u_{is}) \neq 0$ for $s < t$ [correlated with past errors] (intertemporal correlations)

eg) dynamic panel data model

$w_{it} = y_{i,t-1}$. $Cov(y_{it-1}, u_{it}) = 0$

$Cov(y_{it-1}, u_{is}) = 0$ $s > t-1$

but $cov(y_{it-1}, u_{is}) \neq 0$ $s < t$ $(s = t-1, t-2, \ldots$

; weakly exogenous

Case 3 $Cov(w_{it}, u_{it}) \neq 0$, $Cov(w_{it}, u_{is}) = 0$ for $s > t$

$Cov(w_{it}, u_{is}) \neq 0$, $s < t$: [Case 1 + Case 2]

; contemporaneously correlated & weakly exogenous

Case 4 $Cov(w_{it}, u_{is}) \neq 0$, $t \gtreqless s$ (no solution : Forget it)

Case 5 $Cov(w_{it}, u_{is}) = 0$, $t \gtreqless s$ (strictly exog : No problem) more generally,

$$y_{it} = \alpha_i \, (+\rho\, y_{it-1}) + X_{it}\gamma + W_{it}\delta + R_{it}c + e_{it} \qquad (1)$$

dynamic — strictly exog — predetermined — endogenous

i) $E(X_{it}e_{is} = 0)$, for any $s \gtreqless t$: strictly exogenous ((case5))

correlated with past errors ⇒ ii) $E(W_{it}e_{is} = 0)$, for $s \geq t$: predetermined (feedback effect)
(But $E(W_{it}e_{is}) \neq 0$ for $s < t$)
… weakly exogenous (case 2)

Correlated with current & past errors ⇒ iii) $E(R_{it}e_{is} = 0)$, for $s > t$, But $E(R_{it}e_{it}) \neq 0$ $s = t$
$E(R_{it}e_{is}) \neq 0$ $s < t$
… endogenous since $E(R_{it}e_{it}) \neq 0$
⇒ we can have IVs : z_{it}.
(traditional IVs, z_{it}) (case3)

Note ((case2)) y_{it-1} is a part of W_{it} (predetermined variables).

$$(y_{it} - \bar{y}_{i\cdot}) = \rho(\underbrace{y_{it} - \bar{y}_{i-1}}) + \cdots + (\underbrace{e_{it} - \bar{e}_{i\cdot}})$$

These are correlated, and FEs are inconsistent (unless T is big)

$$\left[\begin{array}{l} \bar{y}_{i-1} = \left(\sum_{t=2}^{T} y_{it-1}\right)/(T-1) = \frac{1}{T-1}(y_{i1} + \cdots + y_{it} + \cdots + y_{iT-1}) \\ \bar{e}_{i\cdot} = \left(\sum_{t=2}^{T} e_{it}\right)/(T-1) = \frac{1}{T-1}(e_{i1} + \cdots + e_{it-1} + e_{it} + \cdots + e_{iT-1}) \end{array}\right.$$

⇒ Here, y_{it} is correlated with $(e_{i1} + \cdots + e_{it-1})$, for each t.

Note we need to control for α_i using FE or FD.
FE does not work well in the presence of W_{it} (also y_{it-1})
thus we often use FD to eliminate α_i

Case 1 "Contemporaneous" Correlation

⇒ Are you sure?

$$y_{it} = X_{it}\beta + W_{it}\gamma + \alpha_i + u_{it}$$

; W_{it} affects y_{it} only at time t.

X_{it}: exogenous; W_{it}: endogenous : $IV = Z_{it}$

Solution : 2SLS panel models "xtivreg"

i) FE (WI) estimator

$y_{it}^* = y_{it} - \bar{y}_{i\cdot}$ (or $y_{it} - \bar{y}_{i\cdot} + \bar{y}_{\cdot\cdot}$, where $\bar{y}_{\cdot\cdot}$ is the grand mean)

Also, for X_{it}^*, W_{it}^* and Z_{it}^* (IV)

Then do 2SLS

$$y_{it}^* = X_{it}^*\beta + \underbrace{W_{it}^*}_{IV = Z_{it}^*}\gamma + (\cancel{\alpha_i - \alpha_i}) + u_{it}^*$$

1st Reduced form : W_{it}^* on X_{it}^* and Z_{it}^* ⇒ fitted values $\hat{W}_{it}^*$

2nd $y_{it}^* = X_{it}^*\beta + \hat{W}_{it}^*\gamma + u_{it}^*$ OLS

ii) RE estimator

First, we obtain the within residuals $\tilde{u}_{it}$

then compute $\hat{\theta}_{it}$, and quasi-demeaning : $y_{it}^* = y_{it} - \hat{\theta}_{it}\bar{y}_{i\cdot}$

also for others.

where $\hat{\theta}_{it} = 1 - \dfrac{\hat{\sigma}_u^2}{\hat{\sigma}_{Vi}}$

with $\hat{\sigma}_u^2 = \dfrac{1}{N-n-k+1}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\tilde{u}_{it}^2$

N = total # of obs

n = # of groups

$\hat{\sigma}_\alpha^2 = \dfrac{1}{N}\left[\sum_i^n\sum_t^{T_i}\bar{u}_{it}^2 - (n-k)\hat{\sigma}_u^2\right]$

$\hat{\sigma}_{Vi} = T_i\hat{\sigma}_\alpha^2 + \hat{\sigma}_u^2$

2nd, do 2SLS on y_{it}^* on X_{it}^* and W_{it}^* with IV Z_{it}^*.

iii) FD estimator

$$\Delta y_{it} = \Delta X_{it}\beta + \Delta W_{it}\gamma + \Delta u_{it}$$

- Then, do 2SLS with IV Δz_{it}. But the FD method
- But, the FD method can induce more useful IVs from dynamic relationships.

more IVs = $W_{it-1}, W_{it-2}, \cdots ; \Delta W_{it-1}, \Delta W_{it-2}, \cdots$; also $\Delta X_{it}, X_{it}, X_{it-1}, \cdots ; X_{it+1}, \cdots$

Exercise Stata "xtivreg.do"

Empirical issues Wooldridge, p. 308 - 314

eg1) $\log(wage_{it}) = X_{it}\beta + cigs_{it}\gamma + \alpha_i + u_{it}$ "cigarette smoking"

- $cigs_{it}$ can be correlated with u_{it}. why?
(normal good; consumption ↑ as wage ↑ ; same ind. characteristic

see Levine et al (1997)

- Levine considered the FD method.

: $\Delta \log w_{it} = \Delta X_{it}\beta + \Delta cigs_{it}\gamma + \Delta u_{it}$

IVs for $\Delta cigs_{it}$ can be :

i) entire X_{it} & ΔX_{it} (X_{it} is strictly exogenous)

ii) past smoking $cigs_{i,t-2}$, $cigs_{i,t-3}$ ..

(if $cigs_{it}$ has only a contemporaneous effect on y_{it})

But, this assumption may not be correct.

(One may use $cigs_{i,t-1}$ in X_{it}, instead.)

iii) IVs = cigarette prices, excise tax rates, share of smokers in the population

(But, weak IVs) They avoided this method.

iv) Differencing over sibling's data set

5

```
* xtivreg (Panel IV linear models)

log using xtivreg.log, replace

clear
set memory 40m
set more off

* use abdata.dta, clear
use http://www.stata-press.com/data/r8/nlswork, clear

generate age2 = age^2

tsset idcode year
  * Some commands require tsset.

* FE without endogeneity correction
xtreg ln_w age* tenure not_smsa union south, fe i(idcode)

* FE with endogeneity correction
xtivreg ln_w age* not_smsa (tenure = union south), fe i(idcode)

* RE with endogeneity correction
generate byte black = (race == 2)
xtivreg ln_w age* not_smsa black (tenure = union birth south), re i(idcode)

* FD with endogeneity correction
xtivreg ln_w age* not_smsa black (tenure = union birth south), fd i(idcode)
```

6

```
-------------------------------------------------------------------------------
---
       log:  C:\Documents and Settings\jlee\My
       Documents\Teaching\EC671\xtivreg.l
> og
  log type:  text
 opened on:  24 Aug 2007, 01:40:26

.
. clear

. set memory 40m
(40960k)

. set more off

.
. * use abdata.dta, clear
. use http://www.stata-press.com/data/r8/nlswork, clear
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)

.
. generate age2 = age^2
(24 missing values generated)

.
. tsset idcode year
       panel variable:  idcode, 1 to 5159
        time variable:  year, 68 to 88, but with gaps

.   * Some commands require tsset.
.
. * FE without endogeneity correction
. xtreg ln_w age* tenure not_smsa union south, fe i(idcode)

Fixed-effects (within) regression               Number of obs      =     19007
Group variable (i): idcode                      Number of groups   =      4134

R-sq:  within  = 0.1333                         Obs per group: min =         1
       between = 0.2375                                        avg =       4.6
       overall = 0.2031                                        max =        12

                                                F(6,14867)         =    381.19
corr(u_i, Xb)  = 0.2074                         Prob > F           =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0311984   .0033902     9.20   0.000     .0245533    .0378436
        age2 |  -.0003457   .0000543    -6.37   0.000    -.0004522   -.0002393
      tenure |   .0176205   .0008099    21.76   0.000     .0160331    .0192079
    not_smsa |  -.0972535   .0125377    -7.76   0.000    -.1218289    -.072678
       union |   .0975672   .0069844    13.97   0.000     .0838769    .1112576
       south |  -.0620932    .013327    -4.66   0.000    -.0882158   -.0359706
       _cons |   1.091612   .0523126    20.87   0.000     .9890729    1.194151
-------------+----------------------------------------------------------------
     sigma_u |   .3910683
     sigma_e |  .25545969
         rho |  .70091004   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(4133, 14867) =     8.31         Prob > F = 0.0000

.
. * FE with endogeneity correction
. xtivreg ln_w age* not_smsa (tenure = union south), fe i(idcode)

Fixed-effects (within) IV regression         Number of obs      =      19007
Group variable: idcode                       Number of groups   =       4134
```

7

```
R-sq:  within  =      .                         Obs per group: min =         1
       between = 0.1304                                        avg =       4.6
       overall = 0.0897                                        max =        12

                                                Wald chi2(4)       = 147926.58
corr(u_i, Xb)  = -0.6843                        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      tenure |   .2403531   .0373419     6.44   0.000     .1671643    .3135419
         age |   .0118437   .0090032     1.32   0.188    -.0058023    .0294897
        age2 |  -.0012145   .0001968    -6.17   0.000    -.0016003   -.0008286
    not_smsa |  -.0167178   .0339236    -0.49   0.622    -.0832069    .0497713
       _cons |   1.678287   .1626657    10.32   0.000     1.359468    1.997106
-------------+----------------------------------------------------------------
     sigma_u |  .70661941
     sigma_e |  .63029359
         rho |  .55690561   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F  test that all u_i=0:     F(4133,14869) =     1.44     Prob > F    = 0.0000
------------------------------------------------------------------------------
Instrumented:   tenure
Instruments:    age age2 not_smsa union south

.
. * RE with endogeneity correction
. generate byte black = (race == 2)

. xtivreg ln_w age* not_smsa black (tenure = union birth south), re i(idcode)

G2SLS random-effects IV regression              Number of obs      =     19007
Group variable: idcode                          Number of groups   =      4134

R-sq:  within  = 0.0664                         Obs per group: min =         1
       between = 0.2098                                        avg =       4.6
       overall = 0.1463                                        max =        12

                                                Wald chi2(5)       =   1446.37
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      tenure |   .1391798   .0078756    17.67   0.000      .123744    .1546157
         age |   .0279649   .0054182     5.16   0.000     .0173454    .0385843
        age2 |  -.0008357   .0000871    -9.60   0.000    -.0010063    -.000665
    not_smsa |  -.2235103   .0111371   -20.07   0.000    -.2453386   -.2016821
       black |  -.2078613   .0125803   -16.52   0.000    -.2325183   -.1832044
       _cons |   1.337684   .0844988    15.83   0.000     1.172069    1.503299
-------------+----------------------------------------------------------------
     sigma_u |  .36582493
     sigma_e |  .63031479
         rho |  .25197078   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   tenure
Instruments:    age age2 not_smsa black union birth_yr south

.
. * FD with endogeneity correction
. xtivreg ln_w age* not_smsa black (tenure = union birth south), fd i(idcode)

First-differenced IV regression                 Number of obs      =      5934
Group variable: idcode                          Number of groups   =      3461

R-sq:  within  = 0.1458                         Obs per group: min =         1
       between = 0.6293                                        avg =       4.3
```

```
      overall = 0.0951                                  max =         11

                                                chi2(4)           =       8.52
corr(u_i, Xb)  = -0.4555                        Prob > chi2       =     0.0741

------------------------------------------------------------------------------
d.ln_wage    |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
tenure       |
          D1 |   .1320532   .0755474     1.75   0.080    -.0160171    .2801234
age          |
          D1 |   .0625048   .0267263     2.34   0.019     .0101223    .1148874
age2         |
          D1 |  -.0007977   .0003378    -2.36   0.018    -.0014599   -.0001356
not_smsa     |
          D1 |  -.0610653   .0378044    -1.62   0.106    -.1351607      .01303
black        |
          D1 |  (dropped)
_cons        |  -.0844901   .0631594    -1.34   0.181    -.2082803    .0393001
-------------+----------------------------------------------------------------
     sigma_u |   .5360103
     sigma_e |  .28346454
         rho |  .78144937   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   tenure
Instruments:    age age2 not_smsa black union birth_yr south

.
end of do-file

. exit, clear
```

8

Empirical Homework

EC671 9

1. (Based on Wooldridge book, Introductory Econometrics) In order to estimate the causal effect of prison population increases on crime rates at the state level, Levitt (1996) used instances of prison overcrowding litigation as instruments for the growth in prison population. The equation Levitt estimated is in first differences. We can write an underlying FE model as

$$\log(crime_{it}) = \gamma_t + \alpha_1 \log(prison_{it}) + z_{it}\beta + a_i + u_{it}$$

where γ_t denotes different time intercepts, and crime and prison are measured per 100,000 people. The vector z_{it} includes measures of police per capita, income per capita, unemployment rate, race, and metropolitan and age distributional proportions.

Data: prison.dta from http://www.stata.com/texts/eacsap/

(a) Levitt noted the simultaneity problem between crime rates and prison population, and used 2SLS based on the FD regression, where the instruments for prison are two binary variables, one each for whether a final decision was reached on overcrowding litigation in the current year or in the previous two years. Estimate the FD regression using 2SLS.

(b) One may alternatively use the pooled OLS, fixed effects or random effects models. Estimate the above regression using the pooled OLS, FE and RE estimators (with no endogeneity corrections). Test if the time fixed effects are significant in each of the models. Determine which model is proper among three results.

(c) Using the instruments as in (a), re-estimate the pooled OLS, FE and RE models with endogeneity corrections. Determine which model is proper among three results.

(d) Using the best model of your choice among three results in (c), test a relevant hypothesis for endogeneity.

(e) Using the best model of your choice among three results in (c), determine if your instruments are proper.

Hausman & Taylor model (case 5, but time-invariant variables are included.)

$$y_{it} = \underset{\text{exog}}{X_{1it}\beta_1} + \underset{\substack{\text{endog}\\ \text{(corr with } \alpha_i \\ \text{but not with } \varepsilon_{it})}}{X_{2it}\beta_2} + \underset{\text{exog}}{z_{1i}\delta_1} + \underset{\substack{\text{endo}\\ \text{(corr with } \alpha_i \\ \text{but not with } \varepsilon_{it})}}{z_{2i}\delta_2} + \mu_i + \varepsilon_{it}$$

- X_{2it} (time-varying), Z_{2i} (time invariant) $\Rightarrow$ corr with μ_i but not with ε_{it}, thus RE cannot be used
- FE cannot be used since it does not allow for z_{1i} & z_{2i} (time-invariant).

1st obtain $\hat{\beta}_1$ & $\hat{\beta}_2$ by FE, and the FE-residuals, called $\hat{d}_{it}$

2nd Regress $\hat{d}_{it}$ on z_{1i} and $\underbrace{Z_{2i}}_{\text{endo}}$ using X_{1it} & z_{1i} as IVs. exog. in the system

and obtain $\hat{\delta}_{1IV}$ & $\hat{\delta}_{2IV}$.

(# of var. in X_{1it} > # of var. in z_{2i})

3rd Then obtain within & overall residuals using $\hat{\beta}_{1FE}, \hat{\beta}_{2FE}, \hat{\delta}_{1IV}, \hat{\delta}_{2IV}$ and compute

$$\hat{\theta}_i = 1 - \left(\frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 + T_i\hat{\sigma}_\alpha^2}\right)^{\frac{1}{2}}$$

where $\hat{\sigma}_\alpha^2 = (s^2 - \hat{\sigma}_\varepsilon^2)/\bar{T}$, $\bar{T} = \dfrac{n}{\sum_{i=1}^{n}\frac{1}{T_i}}$

$$s^2 = \frac{1}{N}\sum_i^n\sum_t^{T_i}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}\hat{e}_{it}\right)^2$$

$$\hat{e}_{it} = y_{it} - X_{1it}\hat{\beta}_{1FE} - X_{2it}\hat{\beta}_{2FE} - z_{1i}\hat{\delta}_{1IV} - z_{2i}\hat{\delta}_{2IV}$$

$$\sigma_\varepsilon^2 = \frac{1}{N-n}\sum_1^N\sum_1^T\hat{d}_{it}^2$$

4th GLS transformation using $\hat{\theta}_i$, as in usual RE models.

In summary,

$$y_{it} = \underset{\substack{\downarrow \\ IV = X_{1it}}}{X_{1it}\beta_1} + \underbrace{X_{2it}\beta_2}_{\text{endo}} + \underset{\substack{\downarrow \\ IV = Z_{1i}}}{Z_{1i}\delta_1} + \underbrace{Z_{2i}\delta_2}_{\text{endo}} + \mu_i + \varepsilon_{it}$$

$IV = X_{2it} - \overline{X}_{2i\cdot}$ (as in FE)

$IV = \overline{X}_{1\cdot}$ (the required identification condition is that # of regressors in X_{1it} $\geq$ # of time invariant endogenous regressors Z_{2i}.)

$\Rightarrow$ the points are

i) Time averages of time-varying regressors that are uncorrelated with μ_i are used as IVs for Z_{2i}.

ii) the time demeaned variables of X_{2it} are used as IVs for X_{2it}.

- Exercise : stata. "XTHTAYLOR.DO."
- extensions of H-T estimator

i) Amemiya & MaCurdy (1986)

Use time-invariant instruments

$X_{1,i1} - \overline{X}_{1i}$ up to $X_{1,iT} - \overline{X}_{1i}$

based on the assumption

$E[(X_{1it} - \overline{X}_{1i})\mu_i] = 0$ for each t.

ii) Breusch, Mizon & Schmidt (1989)

Additional IVs are possible

$X_{2i,1} - \overline{X}_{2i}$ up to $X_{2,iT} - \overline{X}_{2i}$

iii) Dynamic panel data models (later) essentially extend these ideas further, and use more moment conditions in the GMM framework.

Hausman & Taylor : simple estimation strategy

ISAW (1999) see Book p327

$$y_{it} = x_{it}\beta + z_i\gamma + \alpha_i + u_{it}$$

$$x_{it} = (x_{1it}, x_{2it})', \quad z_i = (z_{1i}, z_{2i})'$$

1st Pooled 2SLS using IVs

$$(x_{it} - \bar{x}_{i\cdot}), (z_{1i}, \bar{x}_{1i\cdot}) \quad \text{for } x_{it}, z_i$$

That is, (regressors) (IVs)

$x_{1it} \Rightarrow x_{1it} - \bar{x}_{1i\cdot}$ itself but use means

$\boxed{x_{2it}} \Rightarrow x_{2it} - \bar{x}_{2i\cdot}$ using means $\Leftarrow$ key

$z_{1i} \Rightarrow z_{1i}$ itself

$\boxed{z_{2i}} \Rightarrow \bar{x}_{1i\cdot}$ $\Leftarrow$ key

2nd Using pooled 2SLS residuals, obtain $\hat{\lambda}$ and perform quasi-time demeaning for all variables

$$(\tilde{y}_{it} = y_{it} - \hat{\lambda}\bar{y}_{i\cdot} \text{ and so on.})$$

3rd Pooled 2SLS using quasi-time demeaned variables

$\tilde{y}_{it}$ on $\tilde{x}_{1it}, \tilde{x}_{2it}, \tilde{z}_{1i}, \tilde{z}_{2i}$ using IVs

IVs $= \tilde{x}_{1i}, \tilde{x}_{2it}, \tilde{z}_{1i}$ and $\bar{x}_{1i\cdot}$

... Generalized IV estimator

Hausman & Taylor (alternative estimator) 2nd step procedure using transformed data.

$$\tilde{y}_{it} = \tilde{X}_{1it}\beta_1 + \tilde{X}_{2it}\beta_2 + \tilde{z}_{1i}\delta_1 + \tilde{z}_{2i}\delta_2 + \tilde{\varepsilon}_{it}$$

where $\hat{y}_{it} = y_{it} - \hat{\lambda}\bar{y}_{i\cdot}$, $\quad \tilde{z}_{1i} = z_{1i} - \hat{\lambda} z_{1i}$ or $\bar{z}_{1i}$

$\hat{X}_{1it} = X_{1it} - \hat{\lambda}\bar{X}_{1i\cdot}$, $\quad \hat{z}_{2i} = z_{2i} - \hat{\lambda} z_{2i}$ or $\bar{z}_{2i}$

$\hat{X}_{2it} = X_{2it} - \hat{\lambda}\bar{X}_{2i\cdot}$

Use IVs

$$\left\{\begin{array}{lll} (X_{1it} - \bar{X}_{1i\cdot}) & \text{for} & \tilde{X}_{1it} \\ (X_{2it} - \bar{X}_{2i\cdot}) & \text{for} & \tilde{X}_{2it} \\ z_{1i} & \text{for} & \tilde{z}_{1i} \quad (\text{itself}) \\ \bar{X}_{1i\cdot} & \text{for} & \tilde{z}_{2i} \quad \Leftarrow \text{key} \end{array}\right.$$

Note Amemiya & McCurdy (1986) use

$(X_{1it} - \bar{X}_{1i\cdot})$, $(X_{2it} - \bar{X}_{2i\cdot})$, z_{1i} and $(\underbrace{X_{1i1}, X_{1i2}, \cdots X_{1iT}}_{\text{entire history of } X_{1i} \text{ rather than } \bar{X}_{1i\cdot}})$

Example Dep = log-wage

- X_{1it} = (experience, bad health, unemployment last year)
  ... time varying exogenous
- X_{2it} = (whatever)
  ... time varying endogenous
- z_{1i} = (race, union status)
  ... time invariant exogenous
- z_{2i} = (schooling) .. key variable of interest.
  ... time invariant endogenous (to α_i)

14

```
* Hausman and Taylor Method

clear
set memory 40m
set more off

log using xthtaylor.log, replace

webuse xthtaylor1.dta,clear
save xthtaylor1.dta, replace

* use http://www.stata-press.com/data/r8/xthtaylor1, clear

use xthtaylor1.dta, clear

** Hausman-taylor estimator with only endogenous variables
correlate ui z1 z2 x1a x1b x2 eit
xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id)
xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) amacurdy
xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) small
xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) amacurdy small

** Hausman-taylor estimator with constant variables
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) constant(z1 z2 ui) i(id)
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) constant(z1 z2 ui) i(id) t(t)
amacurdy
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) constant(z1 z2 ui) i(id) t(t)
small

** Hausman-taylor estimator with varying variables
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) varying(x2 x1a x1b) i(id)
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) varying(x2 x1a x1b) i(id) t(t)
amacurdy
xthtaylor yit x1a x1b x2 z1 z2 ui, endog(x2 z2) varying(x2 x1a x1b) i(id) t(t)
small

clear
set memory 40m
set more off
set matsize 300

webuse psidextract.dta,clear
save psidextract.dta, replace

* use http://www.stata-press.com/data/r8/psidextract, clear

use psidextract.dta, clear
iis id
tis t

xtsum exp exp2 wks ms union, i(id)

** Hausman-taylor estimator with only endogenous variables
correlate fem blk occ south smsa ind ed

xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed)
xthtaylor lwage occ south smsa ind exp* wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) amacurdy
xthtaylor lwage occ south smsa ind exp* wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) small
xthtaylor lwage occ south smsa ind exp* wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) amacurdy small

** Hausman-taylor estimator with constant variables
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ constant(fem blk ed)
```

15

```
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ constant(fem blk ed) amacurdy
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ constant(fem blk ed) small

** Hausman-taylor estimator with varying variables
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ varying(ms exp* occ south smsa ind wks union)
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ varying(ms exp* occ south smsa ind wks union) amacurdy
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) /*
*/ varying(ms exp* occ south smsa ind wks union) small


** Hausman-taylor estimator with constant variables
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) constant(fem blk ed)
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) constant(fem blk ed) amacurdy
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) constant(fem blk ed) small

** Hausman-taylor estimator with varying variables
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) varying(ms exp* occ south smsa ind wks union)
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) varying(ms exp* occ south smsa ind wks union) amacurdy
xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, endog(exp
exp2 wks ms union ed) varying(ms exp* occ south smsa ind wks union) small
```

16

```
--------------------------------------------------------------------------------
---
       log:  C:\Documents and Settings\jlee\My
       Documents\Teaching\EC671\xthtaylor
> .log
  log type:  text
 opened on:  24 Aug 2007, 01:49:20

.
. webuse xthtaylor1.dta,clear

. save xthtaylor1.dta, replace
file xthtaylor1.dta saved

.
. * use http://www.stata-press.com/data/r8/xthtaylor1, clear
.
. use xthtaylor1.dta, clear

.
. ** Hausman-taylor estimator with only endogenous variables
. correlate ui z1 z2 x1a x1b x2 eit
(obs=10000)

             |       ui       z1       z2      x1a      x1b       x2      eit
-------------+---------------------------------------------------------------
          ui |   1.0000
          z1 |   0.0268   1.0000
          z2 |   0.8777   0.0286   1.0000
         x1a |  -0.0145   0.0065  -0.0034   1.0000
         x1b |   0.0026   0.0079   0.0038  -0.0030   1.0000
          x2 |   0.8765   0.0191   0.7671  -0.0192   0.0037   1.0000
         eit |   0.0060  -0.0198   0.0123  -0.0100  -0.0138   0.0092   1.0000


. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id)

Hausman-Taylor estimation                       Number of obs      =     10000
Group variable (i): id                          Number of groups   =      1000

                                                Obs per group: min =        10
                                                               avg =        10
                                                               max =        10

Random effects u_i ~ i.i.d.                     Wald chi2(5)       =  24172.91
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         yit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         x1a |   2.959736   .0330233    89.63   0.000     2.895011     3.02446
         x1b |   2.953891   .0333051    88.69   0.000     2.888614    3.019168
TVendogenous |
          x2 |   3.022685    .033085    91.36   0.000     2.957839     3.08753
TIexogenous  |
          z1 |   2.709179    .587031     4.62   0.000      1.55862    3.859739
TIendogenous |
          z2 |   9.525973   8.572966     1.11   0.266    -7.276732    26.32868
             |
       _cons |   2.837072   .4276595     6.63   0.000     1.998875    3.675269
-------------+----------------------------------------------------------------
     sigma_u |   8.729479
     sigma_e |  3.1657492
         rho |  .88377062   (fraction of variance due to u_i)
------------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.
```

17

```
. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) amacurdy

Amemiya-MaCurdy estimation                  Number of obs      =     10000
Group variable (i): id                      Number of groups   =      1000

                                            Obs per group: min =        10
                                                           avg =        10
                                                           max =        10

Random effects u_i ~ i.i.d.                 Wald chi2(5)       =  26543.37
                                            Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         yit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         x1a |   2.959344   .0315431    93.82   0.000      2.89752    3.021167
         x1b |   2.954281   .0318127    92.86   0.000     2.891929    3.016633
TVendogenous |
          x2 |   3.023333   .0316108    95.64   0.000     2.961377    3.085289
TIexogenous  |
          z1 |   2.934426    .275198    10.66   0.000     2.395048    3.473804
TIendogenous |
          z2 |   5.780192    1.02449     5.64   0.000     3.772229    7.788155
             |
       _cons |   2.979106   .2683036    11.10   0.000     2.453241    3.504971
-------------+----------------------------------------------------------------
     sigma_u |   8.729479
     sigma_e |  3.1657492
         rho |  .88377062   (fraction of variance due to u_i)
------------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.

. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) small

Hausman-Taylor estimation                   Number of obs      =     10000
Group variable (i): id                      Number of groups   =      1000

                                            Obs per group: min =        10
                                                           avg =        10
                                                           max =        10

Random effects u_i ~ i.i.d.                 Wald chi2(5)       =  24172.91
                                            Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         yit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
TVexogenous  |
         x1a |   2.959736   .0330233    89.63   0.000     2.895003    3.024468
         x1b |   2.953891   .0333051    88.69   0.000     2.888606    3.019176
TVendogenous |
          x2 |   3.022685    .033085    91.36   0.000     2.957832    3.087538
TIexogenous  |
          z1 |   2.709179    .587031     4.62   0.000      1.55848    3.859878
TIendogenous |
          z2 |   9.525973   8.572966     1.11   0.267    -7.278766    26.33071
             |
       _cons |   2.837072   .4276595     6.63   0.000     1.998774    3.675371
-------------+----------------------------------------------------------------
     sigma_u |   8.729479
     sigma_e |  3.1657492
         rho |  .88377062   (fraction of variance due to u_i)
------------------------------------------------------------------------------
note:  TV refers to time-varying; TI refers to time-invariant.

. xthtaylor yit x1a x1b x2 z1 z2, endog(x2 z2) i(id) t(t) amacurdy small
```

[Case 2] Weakly exogenous (inter-temporal correlation)

; $Cov(w_{it}, u_{is}) \neq 0$, $s < t$, but $Cov(w_{it}, u_{is}) = 0$, $s > t$

Examples: strict exogeneity fails due to feedback effects.

eg1) $Patents_{it} = \alpha_i + \gamma_t + z_{it}\beta + \delta R\&D_{it} + \cdots + u_{it}$

(lecture 5, p.14) (also p.13)

$\Rightarrow E(R\&D_{is}\, u_{it}) \neq 0,\ s \geq t$

.. shocks to patents today $(u_{it};\ y_{it})$ affect future R&D spendings

eg2) $(\text{Condom Sales})_{it} = \alpha_i + \gamma_t + z_{it}\beta + \delta \begin{pmatrix}\text{HIV infection}\\ \text{rates}\end{pmatrix}_{it} + u_{it}$

$\Rightarrow E(HIV_{is}\ u_{it}) \neq 0,\ s \geq t$

... Past condom sales (u_{it}) affect HIV infection rates in the future. $\Rightarrow (\text{HIV rates})_{it} = \cdots + \delta_i^* \begin{pmatrix}\text{Condom}\\ \text{Sales}\end{pmatrix}_{it-1} + \cdots$ $(\delta_i^* \neq 0)$

eg3) $\begin{pmatrix}\text{Percentage of flights}\\ \text{canceled}\end{pmatrix}_{it} = \alpha_i + \gamma_t + z_{it}\beta + \delta \begin{pmatrix}\text{Airline}\\ \text{profits}\end{pmatrix}_{it} + u_{it}$

$\Rightarrow E(Profit_{is}\ u_{it}) \neq 0,\ s \geq t$

... Poor performance (u_{it}) will affect profits in subsequent years.

problem All three methods (FE, RE, FD) can lead to biased estimators.
(or inconsistent)

Note: How about using lagged variables? (tempting, but.)

(1) $y_{it} = \alpha_i + \gamma_t + z_{it}\beta + \delta W_{it} + e_{it}$

not strictly exogenous

(2) $y_{it} = \alpha_i + \gamma_t + z_{it}\beta + \delta^* W_{i,t-1} + u_{it}$

$\Rightarrow$ one may be tempted use $W_{i,t-1}$ instead of W_{it}.

Does it resolve the issue?

$\Rightarrow$ The answer is, not really!

eg) $(\text{Condom Sales})_{it} = \cdots + \delta \underbrace{(\text{HIV rates})_{i,t-1}} + u_{it}$

- It looks reasonable since $\text{HIV}_{i,t-1}$ can affect Condom Sales_{it}, but NOT vice versa.
- However, $(\text{HIV Rates})_{i,t-1}$ is still NOT strictly exogenous.

$$\begin{pmatrix}\text{HIV}\\ \text{Rates}\end{pmatrix}_{i,t-1} = \cdots + c_1\begin{pmatrix}\text{Condom}\\ \text{Sales}\end{pmatrix}_{i,t-1} + c_2\begin{pmatrix}\text{Condom}\\ \text{Sales}\end{pmatrix}_{i,t-2} + \cdots$$

if $c_2 \neq 0$, $E(W_{i,t-1} u_{i,t-2}) \neq 0$

where $W_{i,t-1}$ = condom $\text{sales}_{i,t-1}$

eg) Similarly, eg3 in the previous page can have the same problem even if $(\text{Airline Profit})_{i,t-1}$ is used; poor airline performance this year can affect profits in subsequent years.

$\Rightarrow$ $E(W_{i,t+s} u_{it}) \neq 0$, $s > 0$

strict exogeneity assumption fails

due to feedback from y_{it} (u_{it}) to $W_{i,t+s}$.

Why bias? How serious?

(i) FE Models

$$y_{it} = \alpha_i + \delta w_{it} + e_{it} \qquad \text{(no } \beta_{it} \text{ for simplicity)}$$

$$\Rightarrow (y_{it} - \bar{y}_{i\cdot}) = \delta(w_{it} - \bar{w}_{i\cdot}) + (e_{it} - \bar{e}_{i\cdot}) \quad : \text{FE}$$

$$\hat{\delta} = \delta + \left[\frac{1}{T}\sum_{t=1}^{T} E(w_{it} - \bar{w}_{i\cdot})'(w_{it} - \bar{w}_{i\cdot})\right]^{-1} \cdot \left[\frac{1}{T}\sum_{t=1}^{T} E(w_{it} - \bar{w}_{i\cdot})'\underbrace{(e_{it} - \bar{e}_{i\cdot})}_{\text{or simply } e_{it}}\right]$$

$$\left[\text{like.. } \hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'e\right]$$

$$= \delta + (\text{something})^{-1} \cdot \left[\frac{1}{T}\sum_{1}^{T} \underbrace{E(w_{it} - \bar{w}_{i\cdot})e_{it}}\right]$$

where

$$E[(w_{it} - \bar{w}_{i\cdot})e_{it}] = -E(\bar{w}_{i\cdot} e_{it}),$$

since $E(w_{it}e_{it}) = 0$

$$= \delta + (\text{something})^{-1}\left[-\frac{1}{T}\sum_{1}^{T} E(\bar{w}_{i\cdot} e_{it})\right]$$

$$= \delta + (\text{something})^{-1}\left[- E(\bar{w}_{i\cdot} \bar{e}_{i\cdot})\right]$$

this term includes :

$$\begin{cases} E(w_{i,t+s}\, e_{it}) \neq 0, & s > 0 \\ E(w_{it}\, e_{i,t+s}) \underset{\text{(usually)}}{=} 0, & s > 0 \end{cases}$$

the first term introduces bias!

$$\text{Bias term} = (\text{something})^{-1}\left[-E(\bar{w}_{i\cdot}\bar{e}_{i\cdot})\right] = O(T^{-1})$$

$$= \frac{1}{T}(\text{something} * \text{something}) \to 0 \text{ if } T \text{ is big.}$$

Thus, the bias occurs when T is small.
(in panel data models, T is small!)

Note Dynamic panel data models

$$y_{it} = \alpha_i + \underline{\rho y_{it-1}} + \delta \underline{w_{it}} + z_{it}\beta + e_{it}$$

Not strictly exog.

think: $\boxed{E[\bar{y}_{i0}\, \bar{e}_{i0}] \neq 0}$

Bias $= \frac{1}{T}(\text{something}) = O(T^{-1})$: Hsiao (1986)

(bias = big if $\rho \simeq 1.0$)

; others showed that bias $= O(N^{-1}T^{-\frac{3}{2}})$

implying bias $= \frac{1}{NT^{3/2}}(\text{something})$

Thus, bias depends on both N & T.

$\Rightarrow$ intuitively, the same problem occurs for RE models (quasi-demeaning!)

(ii) FD Models

$$\Delta y_{it} = (\alpha_i - \alpha_i) + \delta \Delta w_{it} + \Delta e_{it}$$

$$\Rightarrow \hat{\delta} = \delta + \left[\frac{1}{T}\sum_1^T E(\Delta w_{it}'\Delta w_{it})\right]^{-1}\left[\frac{1}{T}\sum_1^T E(\Delta w_{it}'\Delta e_{it})\right]$$

more serious!

where

$$E(\Delta w_{it}\Delta e_{it}) = E(w_{it}'e_{it}) + E(w_{it-1}e_{it-1}) - E(w_{it-1}e_{it}) - E(w_{it}'e_{it-1})$$

$$= -E(w_{it}'e_{it-1}) \quad \text{if other terms are zero.}$$

$$= \delta + (\text{something})^{-1}(\text{something})$$

Bias $= O(1)$; does not depend on T.

$\Rightarrow$ Even large samples do not lead to consistent estimators.

points when the strict exogeneity fails.

i) using usual FE, RE & FD estimators can lead to bias.

ii) Bias may not pose a problem if T is big (or N, T are big)

iii) FD estimators can be worse. why?

⇒ However, the suggested solutions normally are based on FD estimators. (more on this, later)

$\boxed{\text{Alternative assumption}}$ (required to find solutions)

"Sequentially exogeneity assumption"

For $y_{it} = \alpha_i + X_{it}\beta + u_{it}$,

$$\boxed{E(u_{it} \mid x_{it}, x_{it-1}, \cdots, x_{i1}, \alpha_i) = 0}$$

or equivalently

$$E(y_{it} \mid x_{it}, x_{it-1}, \cdots, x_{i1}, \alpha_i) = E(y_{it} \mid x_{it}, \alpha_i)$$

$$= \alpha_i + X_{it}\beta$$

.. "when x_{it} and c_i are controlled for, no past values of x_{it} affect the expected value of y_{it}."

; "Regressors are independent of future errors in the presence of y_{it-1} and α_i"

Note

i) The above assumption is different from

$$E(y_{it} \mid x_{it}, \ldots x_{i1}) = E(y_{it} \mid x_{it}) = x_{it}\beta$$

$\because$ α_i is missing here.

ii) The above assumption does not require

$$E(x_{i,t+s}\, u_{it}) = 0$$

Thus, feedback effects from y_{it} on future values of x_{it} are allowed.

iii) x_{it} can include y_{it-1}, but not necessarily. If y_{it-1} is included, it is a dynamic panel model

Estimation under the sequentially exogeneity assumption

1st Remove α_i using either FE or FD

2nd Apply IV estimation using IVs

IVs using FE = strictly exogenous variables, if any
IVs using FD = lagged variables or lagged FDs.

Note IVs using FD are easily found; sometimes, there are too many IVs when the FD method is adopted. However, it is usually difficult to find strictly exogenous IVs when the FE method is used

<u>Example 1</u> $\quad y_{it} = \alpha_i + X_{it}\beta + u_{it}$: sequentially exogenous variables

$$\Rightarrow \boxed{\Delta y_{it} = \Delta X_{it}\beta + \Delta u_{it}} \quad \cdots (1)$$

Now, the sequentially exogeneity assumption implies

$$E(X_{i,t-s}u_{it}) = 0 \ , \quad s \geq 0 \qquad (A1)$$

which also implies

$$E(X_{it-s}\Delta u_{it}) = 0 \ , \quad s \geq 1 \qquad (A2)$$

or

$$E(\Delta X_{it-s}\Delta u_{it}) = 0 \ , \quad s \geq 2 \qquad (A3)$$

For example,

(A2) $\Rightarrow$ $\begin{cases} X_{i,t-1} \text{ is an IV for } \Delta X_{it}. \\ X_{i,t-2},\ X_{i,t-3},\ \cdots,\ X_{i2},\ X_{i1} \text{ are also IVs.} \end{cases}$

since $E(X_{it-1}\Delta u_{it}) = E(X_{it-1}u_{it} - X_{it-1}u_{i,t-1}) = 0$

$E(X_{it-2}\Delta u_{it}) = E(X_{it-2}u_{it} - X_{it-2}u_{it-1}) = 0$

$\vdots$

(A3) $\Rightarrow$ $\begin{cases} \Delta X_{i,t-1} \text{ is an IV for } \Delta X_{it} \\ \Delta X_{i,t-2},\ \Delta X_{it-3},\ \cdots,\ \Delta X_{i2} \text{ are also IVs.} \end{cases}$

since $E(\Delta X_{it-1}\Delta u_{it}) = E(X_{i,t-1} - X_{i,t-2})(u_{it} - u_{it-1})$

$= 0$

; show $E(\Delta X_{it-2}\Delta u_{it}) = 0$

$\vdots$

Thus, there are many different IV estimations.

$$\Delta y_{it} = \Delta X_{it}\beta + \Delta u_{it}$$

$$\downarrow$$

$$IV = \begin{cases} (X_{i,t-1}) \\ (\Delta X_{i,t-1}) \\ (X_{i,t-1},\ \Delta X_{i,t-1}) \\ \vdots \end{cases}$$

and Do 2SLS!

IMPORTANT EXAMPLE (Dynamic panel data model)

$$y_{it} = \alpha_i + \rho y_{i,t-1} + u_{it}$$

This is a special case of (1) when $x_{it} = y_{i,t-1}$.

$$\Rightarrow \Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta u_{it}, \quad t \geq 2 \qquad \text{---} (2)$$

◎ Anderson & Hsiao (1982) suggested IV estimation with IVs $y_{i,t-2}$ or $\Delta y_{i,t-2}$ or both.

$\Rightarrow$ this is a very popular method!

Questions

i) Is $\Delta y_{i,t-2}$ a good IV for $\Delta y_{i,t-1}$ in (2)?

$$\begin{cases} E(\Delta y_{i,t-2}\Delta u_{it}) = 0 & \text{under the seq. exog. assumption} \\ E(\Delta y_{i,t-2}\Delta y_{i,t-1}) \neq 0 & \text{not weakly correlated} \end{cases}$$

ii) Anderson & Hsiao (1982) found that using $\Delta y_{i,t-2}$ has a singularity point and can leads to a large std errors of estimators.

iii) Is $y_{i,t-2}$ a good IV for $\Delta y_{i,t-1}$ in (2)?

$$\begin{cases} E(y_{i,t-2}\Delta u_{i,t}) = 0 & \text{under the seq exg. asmptn.} \\ E(y_{i,t-2}\Delta y_{i,t-1}) \neq 0 & \text{But this correlation can be } \underline{\text{weak}}. \end{cases}$$

iv) Using more IVs can lead to more efficient estimators. Then should we use all possible IVs?

The answer is mixed, yes or no.

- yes, up to a certain point.
- NO, when there are too many IVs.
- Yes, if many weak IVs can be helpful. ... Han & Phillips (2006)

◎ Arellano and BOND (1991) GMM estimator

: so, so popular!!

They suggest using all possible level IVs.

- 3086 citations (Google scholar search (8/23/2007))
- 7486 citations (9/8/2010)

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta u_{it}$$

$\Rightarrow$ IVs for $\Delta y_{i,t-1}$ are $y_{i,t-2}, y_{i,t-3}, \cdots y_{i1}$

$t=3$ y_{i1}

$t=4$ y_{i1}, y_{i2}

$t=5$ y_{i1}, y_{i2}, y_{i3}

$\vdots$

$t=T$ $y_{i1}, y_{i2}, y_{i3}, \cdots, y_{i,T-2}$

Let

$$W_i = \begin{bmatrix} y_{i1} & & & \\ 0 & (y_{i1}, y_{i2}) & & 0 \\ & 0 & \cdots & (y_{i1}, y_{i2}, \dots, y_{i,T-2}) \end{bmatrix}$$

Then $E[W_i' \Delta u_i] = 0$: orthogonality condition

Letting $W = [W_1', \dots, W_N']'$, we use GLS-like transformation:

$$W' \Delta y = W'(\Delta y_{-1}) \rho + W' \Delta u$$

Note It can be shown that

$$E(\Delta u_i \Delta u_i') = \sigma_u^2 (I_N \otimes G_T) \overset{\text{let}}{\equiv} \Sigma$$

where $G_T = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & -1 & \\ & & 2 & -1 \\ 0 & & -1 & 2 \end{bmatrix}$

$$\Rightarrow \hat{\rho}_1 = \left[(W' \Delta y_{-1})' [W' \Sigma W]^{-1} (W' \Delta y_{-1}) \right]^{-1} (W' \Delta y_{-1})' (W' \Sigma W)^{-1} W' \Delta y \quad \text{---} (*)$$

... like: $(X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$ = GLS estimator

This is the 1st step estimator, using G_T.

Since Σ can be estimated better by using the residuals, we can use $\Delta \hat{u}_i$, the residual from (2) using $\hat{\rho}_1$:

$$\hat{\Sigma} = \sum_{i=1}^{N} W_i' (\Delta \hat{u}_i)(\Delta \hat{u}_i)' W_i \quad \text{instead of } \Sigma \text{ in } (*)$$

$$\Rightarrow \hat{\rho}_2 = [(W' \Delta y_{-1})' \hat{\Sigma}^{-1} (W' \Delta y_{-1})]^{-1} (W' \Delta y_{-1})' \hat{\Sigma}^{-1} W' \Delta y$$

This is the 2nd step GMM estimator using $\hat{\Sigma}$.

stata: (dyn_panel_example.dta) "xtabond"

Review using dynamic panel data models

$(\, y_{it} = \alpha_i + \rho\, y_{it-1} + z_{it}\gamma + u_{it} \,)$

(i) Will OLS, FE, RE and FD lead to biased estimators?

yes. why? Explain why!

No. when? $\Rightarrow$ If T is big (or N or T is big)

(ii) Why are they biased? (if T is not big)

① OLS is inconsistent. why?

: y_{it-1} is correlated with e_{it}

$$\begin{cases} y_{it} = c + \rho\, y_{it-1} + z_{it} + e_{it}, & e_{it} = \alpha_i + u_{it} \\ y_{it-1} = \quad \cdots \quad + e_{it-1}, & e_{it-1} = \alpha_i + u_{it-1} \end{cases}$$

② FE uses time-demeaned data

mean of $y_{i,t-1}$

$(\, y_{it-1} - \bar{y}_{-1\,i\cdot} \,)$ is correlated with $(\, u_{it} - \bar{u}_{i\cdot} \,)$

since y_{it-1} is correlated with $\bar{u}_{i\cdot}$

($\bar{u}_{i\cdot}$ includes $u_{i,t-s}$, $s \geq 2$)

[eg $E\, y_{it-1} u_{it-2} \neq 0$]

Bias $= O(N^{-1} T^{-3/2})$

③ RE has the same problem

$(\, y_{it-1} - \lambda \bar{y}_{-1\,i\cdot} \,)$ is correlated with $(\, u_{it} - \lambda \bar{u}_{i\cdot} \,)$

④ FD has the same problem; more serious. why?

$(\, y_{it-1} - y_{it-2} \,)$ is correlated with $(\, u_{it} - u_{it-1} \,)$

Note on GMM dynamic panel models

i) the interpretation can be given in the original equation : $y_{it} = \alpha_i + \beta X_{it} + u_{it}$

: relationship bet y_{it} & X_{it} (or y_{it-1})

(but not Δy_{it} and ΔX_{it})

ii) Using all IVs may not be good.

using 5 lags, for instance, can be good enough.

iii) Arellano & Bover (1995) suggested using both level and differenced IVs.

[all values of $y_{i,t-s}$, and Δy_{it-s}, $s \geq 2$]

... System IVs

iv) Ahn & Schmidt (1995) found more moment conditions (orthogonality conditions)

① $E(y_{is} \Delta u_{it}) = 0$, $t = 2, .., T$, $s = 0, .., t-2$: $T(T-1)/2$

... same as Arellano & Bond (1991)

② $E(u_{iT} \Delta u_{it}) = 0$, $t = 2, .., T-1$: $(T-2)$

③ u_{it} is homoskedastic

④ More IVs from strictly exogenous variables

$$y_{it} = \alpha_i + \rho y_{it-1} + Z_{it}\beta + u_{it}$$

v) Testing $\rho = 1$ (unit root) is beyond the scope.

$\Rightarrow$ panel unit root tests in EC672.

```
log using dyn_panel_example.log, replace

clear
set memory 40m
set more off

set matsize 350

use "al_tax_std_updated.dta", clear
tsset state year

** Anderson-Hsiao FD with IV = y(t-2)

xtivreg lnallgon lnallgonlag percinc percisq  maxafdc teenperc (l.lnallgon =
l2.lnallgon), i(state) fd


** Arellano & Bond One Step: maxlags = all

xtabond lnallgon  percinc percisq  maxafdc teenperc, lags(1)


** Arellano & Bond Two Step: maxlags = all

xtabond lnallgon  percinc percisq  maxafdc teenperc, lags(1) twostep
```

31

```
--------------------------------------------------------------------------------
---
       log:  C:\Documents and Settings\jlee\My
       Documents\Teaching\EC671\dyn_panel
> _example.log
  log type:  text
 opened on:  24 Aug 2007, 01:23:44

.
. clear

. set memory 40m
(40960k)

. set more off

.
. set matsize 350

.
. use "al_tax_std_updated.dta", clear

. tsset state year
       panel variable:  state, 1 to 56
        time variable:  year, 75 to 95

.
. ** Anderson-Hsiao FD with IV = y(t-2)

.
. xtivreg lnallgon lnallgonlag percinc percisq  maxafdc teenperc (l.lnallgon =
> l2.lnallgon), i(state) fd

First-differenced IV regression                 Number of obs      =       915
Group variable: state                           Number of groups   =        51

R-sq:  within  = 0.0586                         Obs per group: min =        16
       between = 0.0006                                        avg =      17.9
       overall = 0.4588                                        max =        18

                                                chi2(5)            =     14.26
corr(u_i, Xb)  = 0.7094                         Prob > chi2        =    0.0065

------------------------------------------------------------------------------
d.lnallgon   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnallgon     |
          LD |   .0879599   .0335083     2.63   0.009       .022285    .1536349
lnallgonlag  |
          D1 |  (dropped)
percinc      |
          D1 |   -.000041   .0000205    -2.00   0.045     -.0000813   -8.27e-07
percisq      |
          D1 |   1.22e-09   5.92e-10     2.06   0.040      5.79e-11    2.38e-09
maxafdc      |
          D1 |   .0000309   .0001918     0.16   0.872     -.0003452    .0004069
teenperc     |
          D1 |  -.0169186   .0100724    -1.68   0.093       -.03666    .0028229
_cons        |  -.0849368   .0076813   -11.06   0.000     -.0999919   -.0698817
-------------+----------------------------------------------------------------
     sigma_u |   .7985461
     sigma_e |   .1669532
         rho |  .95811975   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:    L.lnallgon
Instruments:     lnallgonlag percinc percisq maxafdc teenperc L2.lnallgon

.
.
```

32

```
. ** Arellano & Bond One Step: maxlags = all
.
. xtabond lnallgon  percinc percisq  maxafdc teenperc, lags(1)

Arellano-Bond dynamic panel-data estimation     Number of obs      =       966
Group variable (i): state                       Number of groups   =        51

                                                Wald chi2(4)       =   2064.81

Time variable (t): year                         Obs per group: min =        17
                                                               avg =  18.94118
                                                               max =        19

One-step results
------------------------------------------------------------------------------
lnallgon     |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnallgon     |
          LD |    .860628   .0209267    41.13   0.000     .8196124    .9016436
percinc      |
          D1 |  -.0000314   .0000179    -1.76   0.078    -.0000665    3.56e-06
percisq      |
          D1 |   8.59e-10   3.99e-10     2.16   0.031     7.82e-11    1.64e-09
maxafdc      |
          D1 |   -.000075   .0001766    -0.42   0.671    -.0004211    .0002712
teenperc     |
          D1 |   .0047867   .0085186     0.56   0.574    -.0119094    .0214828
_cons        |  -.0181812   .0048158    -3.78   0.000      -.02762   -.0087423
------------------------------------------------------------------------------
Sargan test of over-identifying restrictions:
         chi2(189) =    475.85     Prob > chi2 = 0.0000

Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation    z = -11.51   Pr > z = 0.0000
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
         H0: no autocorrelation    z =  -2.99   Pr > z = 0.0028


.
.
. ** Arellano & Bond Two Step: maxlags = all
.
. xtabond lnallgon  percinc percisq  maxafdc teenperc, lags(1) twostep

Arellano-Bond dynamic panel-data estimation     Number of obs      =       966
Group variable (i): state                       Number of groups   =        51

                                                Wald chi2(4)       =  28087.15

Time variable (t): year                         Obs per group: min =        17
                                                               avg =  18.94118
                                                               max =        19

Two-step results
------------------------------------------------------------------------------
lnallgon     |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lnallgon     |
          LD |   .8593747   .0070278   122.28   0.000     .8456005    .8731488
percinc      |
          D1 |  -.0000294   2.87e-06   -10.22   0.000     -.000035   -.0000237
percisq      |
          D1 |   8.11e-10   8.28e-11     9.79   0.000     6.48e-10    9.73e-10
maxafdc      |
          D1 |  -.0000671   .0000414    -1.62   0.105    -.0001482     .000014
teenperc     |
          D1 |   .0035355   .0018743     1.89   0.059     -.000138    .0072091
_cons        |  -.0186162    .001014   -18.36   0.000    -.0206037   -.0166287
------------------------------------------------------------------------------
```

```
Warning: Arellano and Bond recommend using one-step results for
         inference on coefficients

Sargan test of over-identifying restrictions:
         chi2(189) =    49.92     Prob > chi2 = 1.0000

Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation   z =  -4.07   Pr > z = 0.0000
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
         H0: no autocorrelation   z =  -1.80   Pr > z = 0.0718

.
end of do-file
```

ANOTHER IMPORTANT EXAMPLE (special case of Example 1)

.. also apply to Case 2 & 3

$T=2$. $y_{it} = \alpha_i + X_{it}\beta + u_{it}$, $t=1,2$ $(T=2)$

Examples include [matched samples,
DID estimator (before & after)

$\Delta y_{it} = \Delta X_{it}\beta + \Delta u_{it}$ becomes

$\Rightarrow \Delta y_{i2} = \Delta X_{i2}\beta + \Delta u_{i2}$ with only N obs $\quad$ — $(*)$ $\quad \left(\begin{array}{c} N\cdot(T-1) = N \\ \text{if } T=2 \end{array} \right)$

Question : when X_{it} is not strictly exogenous?

$\Rightarrow E(X_{it+s} u_{it}) \neq 0$

i.e. $E(X_{i2} u_{i1}) \neq 0$ but $E(X_{i2} u_{i2}) = E(X_{i1} u_1)$
$= E(X_{i1} u_2) = 0$

Example) y_{it} = stock returns [$t=1$ before IPO
$t=2$ after IPO

X_{it} = debts

$\begin{pmatrix}\text{stock}\\ \text{returns}\end{pmatrix}_{it} = \alpha_i + \beta(\text{debts})_{it} + (\text{others}) + u_{it}$

consider

$\underset{\substack{[\text{after IPO}]\\ t=2}}{(\text{debts})_{i2}} = c + \delta \underset{\substack{[\text{before IPO}]\\ t=1}}{\begin{pmatrix}\text{stock}\\ \text{returns}\end{pmatrix}_{i1}} + \cdots + u_i$ $\quad$ (N obs)

If $\delta \neq 0$, X_{it} is not strictly exog.
($\delta < 0$, for instance)

Estimation IV for ΔX_{i2}
is X_{i1} under seq. exog. assumption.

But, X_{i1} and ΔX_{i2} can be weakly correlated.
then, we expect <u>large</u> std. errors of the estimator.

Example 2 Both strictly and sequentially exogenous variables

$$y_{it} = \alpha_i + \underbrace{z_{it}\gamma}_{\text{strictly exog}} + \underbrace{w_{it}\delta}_{\text{sequentially exog}} + u_{it}$$

$$\Rightarrow \Delta y_{it} = \underbrace{\Delta z_{it}\gamma}_{\text{no need to use IV}} + \underbrace{\Delta w_{it}}_{\text{IVs for this}}\delta + \Delta u_{it}$$

$= w_{i,t-s}$, $\Delta w_{i,t-s}$, $s \geq 2$ as in case 1

Plus z_{it}, Δz_{it}, $\Delta z_{i,t-s}$ and even $\Delta z_{i,t+s}$ since z_{it} is strictly exog.

points: we could use additional IVs.

Note the dynamic GMM models can include z_{it}

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \boxed{z_{it}\gamma} + u_{it}$$

thus more IVs are possible

Question When the FE transformation is used (instead of FD), we cannot use lagged values of w_{it} as IVs for $(w_{it} - \overline{w_{i\cdot}})$. why?

More on FD method, using IV estimation

$$\Delta y_{it} = (\rho \Delta y_{it-1}) + \Delta X_{it}\gamma + \Delta W_{it}\delta + \Delta R_{it}c + \Delta e_{it} \qquad (2)$$

1) ΔW_{it}: ΔW_{it} & Δe_{it} are correlated.

$\Rightarrow$ IVs $= W_{it-1}, W_{it-2}, \dots, W_{i1}$ (all past values)

check $E(W_{it-1}\Delta e_{it}) = E(W_{it-1}e_{it}) - E(W_{it-1}e_{it-1})$
$= 0 - 0$

Thus, when $W_{it} = y_{it-1}$, we can use

$\Rightarrow$ IVs $= y_{it-2}, y_{it-3}, \dots, y_{i1}$ (all past values)

check $E(y_{it-2}\Delta e_{it}) = E(y_{it-2}e_{it}) - E(W_{it-2}e_{it-1})$
$= 0 - 0$

2) ΔR_{it}: ΔR_{it} & Δe_{it} are correlated.

$\Rightarrow$ IVs $=$ { ΔZ_{it} ($Z_{it-1}, Z_{it-2}, \dots, Z_{i1}$; $\Delta Z_{it-1}, \dots \Delta Z_{i2}$)
$R_{it-2}, R_{it-3}, \dots, R_{i1}$ (excluding R_{it-1})

check $E(R_{it-2}\Delta e_{it}) = E(R_{it-2}e_{it}) + E(R_{it-2}e_{it-1})$
$= 0 + 0$

But, $E(R_{it-1}\Delta e_{it}) = E(R_{it-1}e_{it}) - E(\underbrace{R_{it-1}e_{it-1}}_{\neq 0})$
$= 0 -$ something
$\neq 0$ (Thus R_{it-1} is not a good IV)

3) ΔX_{it}: ΔX_{it} & Δe_{it} are Not correlated since X_{it} are strictly exogenous.

Thus, we do not need IVs for ΔX_{it}.

But, $X_{it+s}, X_{it-s}, X_{it}$ can be used as IVs for ΔW_{it} and ΔR_{it}, if needed.

(Simple case) (no w_{it}, R_{it}, but y_{it-1} is included) (X_{it})

$$y_{it} = \alpha_i + \rho y_{it-1} + X_{it}\gamma + e_{it} \qquad (1)' \text{ model}$$

$$\Rightarrow \Delta y_{it} = \rho \Delta y_{it-1} + \Delta X_{it}\gamma + \Delta e_{it} \qquad (2)' \text{ estimation}$$

(A-H) Anderson-Hsiao (1982) suggested IV estimation using y_{it-2} or Δy_{it-2} or both as IVs for Δy_{it-1}.

i) y_{it} and Δy_{it-2} are uncorrelated with Δe_{it}

ii) " are correlated with Δy_{it-1}.

Note y_{it-2} and Δy_{it-2} are <u>variables</u>.

; we lose the observations for $t=1,2$
(All observations with $t=1,2$ are dropped off.)

Thus, we <u>cannot</u> utilize the IVs that were given in the above (previous page)

<u>Note</u> STATA command

```
ivreg D.y (D.yL1 = yL2) D.(X1 X2 X3 yr*)
```

(D. = difference; L1 = lag 1; L2 = lag 2; D.( ) = difference)

; $\Delta y_{it} = \rho \Delta y_{it-1} + \gamma_1 \Delta X_{1it} + \gamma_2 \Delta X_{2it} + \gamma_3 \Delta X_{3it} + c_1 \Delta year91_{it} + \cdots + c_{10} \Delta year00_{it} + e_{it}$

$\rho \Delta y_{it-1} \downarrow IV = \underline{y_{it-2}}$

$\Downarrow$

$$IV = y_{it-2} \Rightarrow \underset{\sim}{Z} = \begin{bmatrix} \cdot \\ \cdot \\ y_{i1} \\ \vdots \\ y_{iT-2} \end{bmatrix} \begin{array}{l} t=1 \\ t=2 \\ t=3 \\ \vdots \\ t=T \end{array}$$

; In theory, less efficient, but it often gives good & reasonable results, though.

(A better way using IVs) Not losing obs. (Not too many)

i) $IV = y_{it-2} \Rightarrow$

$$Z_i = \begin{bmatrix} \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & 0 \\ y_{i1} & 0 & \cdots & 0 \\ 0 & y_{i2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & & y_{iT-2} \end{bmatrix}$$

: IV for each time period.

(GMM style IV matrix)

(Holtz Eakin et al)

Holtz-Eakin, Newey & Rosen (1988) "VAR-panel models"

they adopted the above "GMM-style" instruments.

note But, this treatment still uses y_{it-2} only as IV for Δy_{it-1}, but using more observations than A-H estimator

ii) $IV = (y_{it-2}, y_{it-3}, \cdots y_{i1})$: All past values

(Arellano & Bond)

$$Z_i = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ y_{i1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & & & \vdots & \vdots & \vdots & \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cdot & \cdot & \cdot & \cdots \\ y_{i1} & 0 & 0 & \cdots \\ y_{i2} & y_{i1} & 0 & \\ y_{i3} & y_{i2} & y_{i1} & \\ \vdots & & & \ddots \end{bmatrix}$$

collapsed form (preferred)

$\Rightarrow$ Using all previous values!

$\Rightarrow$ $E(Z'E) = 0$ where $Z = \begin{bmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_N \end{bmatrix}$ using Z_i

This is the moment condition of the GMM estimation procedure of Arellano & Bond (1991).

(uses all past values + following the form of Z_i (GMM style) of Holtz-Eakin et al (1988).)

$\Rightarrow$ GMM Estimator minimizes $\xi = (Z'\hat{E})' W (Z'\hat{E})$ to estimate the parameters ($\hat{\rho}$, $\hat{\beta}$ ...) $\Rightarrow$ see p.19 Lecture 6 (alternative way).

Note

i) How big is the bias if we still use FE?
(RE or FD, but not IV)
nor GMM

FE:

$$(y_{it} - \bar{y}_{i.}) = \rho(y_{it-1} - \bar{y}_{i-1}) + (e_{it} - \bar{e}_{i.})$$

correlated

If $\rho > 0$, ρ will be underestimated

$$\text{bias} = E(\hat{\rho}) - \rho \to -\frac{1+\rho}{T-1} \text{ as } N \to \infty$$

(eg. $T=10$, $\rho = 0.5$, bias $= -0.167$)

But, when other variables (x_{it}, w_{it}) appear, the direction of bias can be changed.

ii) Autocorrelation check-up in the FD GMM estimation

Even if e_{it} are serially uncorrelated, Δe_{it} & Δe_{it-1} are correlated.

$$E(e_{it} - e_{it-1})(e_{it-1} - e_{it-2}) = -\sigma_e^2 \neq 0$$

But Δe_{it} and Δe_{it-2} are uncorrelated.

Thus, at lag 1, autocorrelation is often found.

$\Rightarrow$ not a problem

at lag 2, autocorrelation should be absent.

$\Rightarrow$ If not, we need to add y_{it-2}, y_{it-3} so on until we do not observe autocorrelation at lag 2.

stata estat abond : for autocorrelation test

iii) One-step vs two-step estimation

- the first step GMM estimation uses I or G matrix as the weighting matrix W.
- Thus, we often prefer the 2nd step results for the estimator.
- However, it is known that the std error of the 2nd step estimator is biased. Then, Arellano & Bond initially suggested using the std. error of the 1st-step estimator, which may not be the best.
- Then, it's is desirable to use the Windmeijer (2005) robust variance (bias-corrected variance)

  Note when the robust option is used, the robust variance is reported in the 2nd step result.

iv) Checking robustness by reducing # of IVs.

- As # of IVs (weak IVs) increases, the GMM estimator can be biased.
- Thus, it is desirable to use parsimonious # of IVs by restricting # of past values.

  rule: keep: # of IVs < # of groups

stata xtabond y w, lags(2) maxldep(3) maxlags(3)

y_{it-1}, y_{it-2} $(p=2)$; use $y_{it-2}, y_{it-3}, y_{it-4}$ as IV (1 lag)

; maxlags: # of lags of predetermined & endogenous variables for use as IVs.

STATA Commands (xtabond2 & others)

Data: abdata.dta

n_{it} = log of employment (n)
w_{it} = log of real wage (w)
k_{it} = log of gross capital stock (k)
ys_{it} = log of industry output (ys)
yr1976 – yr1984

Notation
L.n = $n_{i,t-1}$, nL1 = $n_{i,t-1}$
L2.n = $n_{i,t-2}$
D.n = Δn_{it} , D.(nL2 w) = $[\Delta n_{i,t-2}, \Delta w_{it}]$
L(0/2).(k ys) = $[k_{it}, k_{it-1}, k_{it-2}, ys_{it}, ys_{it-1}, ys_{it-2}]$
L(2/.).n = $[n_{i,t-2}, n_{i,t-3}, \cdots n_{i1}]$

Ex 1) ivreg D.n (D.nL1 = nL2) D.(nL2 w wL1 k kL1 kL2 ys..)

Anderson-Hsiao estimation

$$\Rightarrow \Delta n_{it} = \rho \underbrace{\Delta n_{it-1}}_{\downarrow\ IV = n_{i,t-2}} + \underbrace{\gamma_1 \Delta n_{it-2}} + \gamma_2 \Delta w_{it} + \gamma_3 \Delta w_{it-1} + \gamma_4 \Delta k_{it} + \gamma_5 \Delta k_{it-1} + \gamma_6 \Delta k_{it-2} + \gamma_7 \Delta ys_{it} + \cdots$$

Ex 2) xtabond n $\underbrace{\text{L(0/1).w}}$ $\underbrace{\text{L(0/2).(k ys) yr*}}$, lags(2)

$$\Rightarrow \Delta n_{it} = \rho_1 \Delta n_{it-1} + \underbrace{\rho_2 \Delta n_{it-2}}_{\text{lags(2)}} + \gamma_2 \Delta w_{it} + \gamma_3 \Delta w_{it-1} + \gamma_4 \Delta k_{it} + \gamma_5 \Delta k_{it-1} + \gamma_6 \Delta k_{it-2} + \gamma_7 \Delta ys_{it} + \gamma_8 \Delta ys_{it-1} + \gamma_9 \Delta ys_{it-2} + \cdots$$

IV = all past values of $n_{i,t-2}$

The same results.

Ex 3) xtabond2 n L.n L2.n w L.w L(0/2).(k ys) yr* ,
gmm(L.n) iv(w L.w L(0/2).(k ys) yr*) h(1)
noleveL

gmm(L.n) ; L.n is predetermined to be instrumented
iv(w L.w ...) ; these are exogenous — more, later
nolevel ; not system GMM, but use "difference GMM"

More generally,

$$y_{it} = \alpha_i + \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + x_{it}\gamma + w_{it}\delta + R_{it}c + e_{it}$$

strictly exog — predetermined — endogenous

```
- xtabond y x, lags(2) maxldep(3)     ///
      pre(w, lag(1, 2))               ///
      endogenous(R, lag(0, 2))        ///
      twostep vce(robust)
```

⇒ maxldep(3); use $y_{it-2}, y_{it-3}, y_{it-4}$ as IVs for Δy_{it-1}

pre(w, lag(1, 2)) — "endmaxlags" ⇒ include w_{it-2}, w_{it-3} as IVs for Δw_{it}

"prelags" ⇒ include $\Delta w_{it}, \Delta w_{it-1}$ as regressors

endogenous(R, lag(0, 2)) — include R_{it-2}, R_{it-3} as IVs for ΔR_{it}

— include ΔR_{it} only

Note if we use: pre(w, lag(0, 2)), it means

use w_{it-1}, w_{it-2} as IV for Δw_{it}

Include Δw_{it} only as a regressor.

```
- xtabond2 y L.y L2.y w R, nolevel
      iv(x)          ... strictly exogenous
      gmm(w R)       .. w & R are predetermined or endogenous
      gmm(L.y L2.y)  .. predetermined
  or
      gmm(w, laglimits(1 .))  ... use w_it-1, w_it-2, ... w_i1 as IVs for Δw_it
      gmm(w, laglimits(1 2))  .. use (w_it-1, w_it-2) as IVs for Δw_it
      gmm(R, laglimits(2 3))  .. use (R_it-2, R_it-3) as IVs for ΔR_it
                     lag 2 to lag 3      endogenous
      gmm(L.y L2.y, laglimits(2 4)) .. use (y_it-2, y_it-3, y_it-4) as IV for Δy_it-1
```

43

```
capture log close
log using panel_2_A_Bond.log, replace

set more off
clear all

set memory 30m
set linesize 90

* Data of Arrelano and Bond

use http://www.stata-press.com/data/r7/abdata.dta, clear
save abdata, replace

tsset id year

************ Pooling OLS

regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*
```

n	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nL1	1.044643	.0336647	31.03	0.000	.9785523	1.110734
nL2	-.0765426	.0328437	-2.33	0.020	-.1410214	-.0120639
w	-.5236727	.0487799	-10.74	0.000	-.6194374	-.427908
wL1	.4767538	.0486954	9.79	0.000	.381155	.5723527
k	.3433951	.0255185	13.46	0.000	.2932972	.3934931
kL1	-.2018991	.0400683	-5.04	0.000	-.2805613	-.123237
kL2	-.1156467	.0284922	-4.06	0.000	-.1715826	-.0597107
ys	.4328752	.1226806	3.53	0.000	.1920285	.673722
ysL1	-.7679125	.1658165	-4.63	0.000	-1.093444	-.4423813
ysL2	.3124721	.111457	2.80	0.005	.0936596	.5312846

```
************ LSDV

xi: regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr* i.id

*** FE alternatively,

xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe
```

n	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nL1	.7329476	.039304	18.65	0.000	.6557563	.810139
nL2	-.1394773	.040026	-3.48	0.001	-.2180867	-.0608678
w	-.5597445	.057033	-9.81	0.000	-.6717551	-.4477339
wL1	.3149987	.0609756	5.17	0.000	.1952451	.4347522

44

```
         k |   .3884188   .0309544    12.55   0.000     .3276256    .4492119
       kL1 |  -.0805185   .0384648    -2.09   0.037    -.1560618   -.0049751
       kL2 |  -.0278013   .0328257    -0.85   0.397    -.0922695     .036667
        ys |    .468666   .1231278     3.81   0.000     .2268481    .7104839
      ysL1 |  -.6285587     .15796    -3.98   0.000    -.9387856   -.3183318
      ysL2 |   .0579764   .1345353     0.43   0.667    -.2062454    .3221982
```

```
*********** Anderson-Hsiao

ivreg D.n (D.nL1= nL2) D.(nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr1979 yr1980 yr1981 yr1982 yr1983)

------------------------------------------------------------------------------
         D.n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         nL1 |
         D1. |   2.307626   1.999547     1.15   0.249    -1.619403    6.234655
             |
         nL2 |
         D1. |  -.2240271   .1814343    -1.23   0.217    -.5803566    .1323025
             |
           w |
         D1. |  -.8103626   .2653017    -3.05   0.002    -1.331404   -.2893209
             |


*********** xtabond using a user defined z matrix

forvalues yr=1978/1984 {
 forvalues lag = 2 / `= `yr' - 1976' {
 quietly generate z`yr'L`lag' = L`lag'.n if year == `yr'
 }
}

quietly recode z* (. = 0)
* replacing missing with zero

ivreg D.n  D.(L2.n w L.w k L.k L2.k ys L.ys L2.ys yr1978 yr1979 yr1980 yr1981 yr1982 yr1983) (DL.n = z*), no
* dep variable = Dn(i,t); labor
* lagged dependent variables = DL.n, DL2.n (2 lags)
* ivs for Dn(i,t-1) = y(i,t-k), k = 2, 3,.. (all previous lags)
* other regressors: differences of w(i,t), w(i,t-1),  k(i,t), k(i,t-1), k(i,t-2),  ys(i,t), ys(i,t-1), ys(i,

. sum z*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     z1978L2 |      1031     .0961561    .5065681  -1.443923   4.587607
     z1979L2 |      1031     .1552987    .6284283  -1.951928   4.597128
```

45

```
     z1979L3 |      1031    .0961561    .5065681  -1.443923    4.587607
     z1980L2 |      1031    .1599216    .6364236  -2.009915    4.609312
     ... (omitted)..
-------------+------------------------------------------------------
     z1984L6 |      1031    .0179951    .2193046  -2.009915    3.020425
     z1984L7 |      1031    .0152084    .2088508  -1.951928    3.030134
     z1984L8 |      1031    .0102655    .1425165   -.486133    2.995732


------------------------------------------------------------------------------
         D.n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         LD. |   .2689418   .1466334     1.83   0.067    -.0190402    .5569238
        L2D. |  -.0669834   .0437388    -1.53   0.126    -.1528845    .0189177
             |
           w |
         D1. |  -.5723355   .0581178    -9.85   0.000    -.6864766   -.4581945
         LD. |   .2112242   .1050951     2.01   0.045     .0048217    .4176266
             |
           k |
         D1. |   .3843826     .03236    11.88   0.000     .3208289    .4479363
         LD. |   .0796079   .0545831     1.46   0.145    -.027591     .1868069
        L2D. |   .0231674   .0369709     0.63   0.531    -.0494419    .0957768

******* Getting the same results using xtabond2 with I matrix

xtabond2 n  L.n L2.n  w L.w   L(0/2).(k ys)  yr*, gmm(L.n)  iv(w L.w  L(0/2).(k ys)  yr*)  h(1) nolevel smal
* dep variable = n(i,t); labor .. will be differenced (also all variables below will be differenced)
* lagged dependent variables = L.n L2.n = n(i,t-1), n(i,t-2)
* ivs for Dn(i,t-1) = n(i,t-k), k = 2, 3,.. (all previous lags)
* other regressors: w(i,t), w(i,t-1),  k(i,t), k(i,t-1), k(i,t-2),  ys(i,t), ys(i,t-1), ys(i,t-2), year*

* Note: other regressors (w L.w  L(0/2).(k ys) yr*) will need to be repeated here with "iv" => iv = w L.w  L
*  to denote that they are exogenous; thus no need to instrument them.

* Note: h(1) implies that H = I matrix (homoskedasticity) is used in GMM estimation.

* Note: The above results appear strange; the coefficient of the lagged dependent variable is too low.

Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 41                      Obs per group: min =         4
F(16, 595)    =     42.63                                      avg =      4.36
Prob > F      =     0.000                                      max =         6
------------------------------------------------------------------------------
           n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

46

```
           n |
         L1. |   .2689418   .1466334     1.83   0.067    -.0190402    .5569238
         L2. |  -.0669834   .0437388    -1.53   0.126    -.1528845    .0189177
             |
           w |
         --. |  -.5723355   .0581178    -9.85   0.000    -.6864766   -.4581945
         L1. |   .2112242   .1050951     2.01   0.045     .0048217    .4176266
             |

********* xtabond or xtabond2 using a proper weighting matrix

xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys) yr*, gmm(L.n) iv(w L.w L(0/2).(k ys) yr*) nolevel robust
* without using h(1)

xtabond  n  l(0/1).w   l(0/2).(k ys)  yr* , lags(2) noconstant
xtabond  n  l(0/1).w   l(0/2).(k ys)  yr* , lags(2)
xtabond  n  l(0/1).w   l(0/2).(k ys)  yr*  year, lags(2) noconstant
xtabond  n  l(0/1).w   l(0/2).(k ys)  yr*  year, lags(2)

* Note: These results are the same; regardless of including year and constant.
* Note: These results do not use IVS from the past values of the predetermined or endogenous variables.

Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 41                      Obs per group: min =         4
Wald chi2(16) =    1727.45                                     avg =      4.36
Prob > chi2   =      0.000                                     max =         6
------------------------------------------------------------------------------
             |              Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6862261   .1445943     4.75   0.000     .4028266    .9696257
         L2. |  -.0853582   .0560155    -1.52   0.128    -.1951467    .0244302
             |
           w |
         --. |  -.6078208   .1782055    -3.41   0.001    -.9570972   -.2585445
         L1. |   .3926237   .1679931     2.34   0.019     .0633632    .7218842
             |
           k |
         --. |   .3568456   .0590203     6.05   0.000      .241168    .4725233
         L1. |  -.0580012   .0731797    -0.79   0.428    -.2014308    .0854284
         L2. |  -.0199475   .0327126    -0.61   0.542    -.0840631    .0441681
             |
          ys |
         --. |   .6085073   .1725313     3.53   0.000     .2703522    .9466624
         L1. |  -.7111651   .2317163    -3.07   0.002    -1.165321   -.2570095
         L2. |   .1057969   .1412021     0.75   0.454    -.1709542     .382548
```

47

```
      yr1978 |   .0077033   .0314106     0.25   0.806    -.0538604     .069267
      yr1979 |   .0172578   .0290922     0.59   0.553    -.0397619    .0742775
      yr1980 |   .0297185   .0276617     1.07   0.283    -.0244974    .0839344
      yr1981 |   -.004071   .0298987    -0.14   0.892    -.0626713    .0545293
      yr1982 |  -.0193555   .0228436    -0.85   0.397     -.064128    .0254171
      yr1983 |  -.0136171   .0188263    -0.72   0.469     -.050516    .0232818
------------------------------------------------------------------------------

Instruments for first differences equation
  Standard
    D.(w L.w k L.k L2.k ys L.ys L2.ys yr1976 yr1977 yr1978 yr1979 yr1980
    yr1981 yr1982 yr1983 yr1984)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/.).L.n

------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.60  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.52  Pr > z =  0.606
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(25)   =  67.59  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(25)   =  31.38  Prob > chi2 =  0.177
  (Robust, but can be weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:
  iv(w L.w k L.k L2.k ys L.ys L2.ys yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr198
> 3 yr1984)
    Hansen test excluding group:     chi2(11)   =  12.01  Prob > chi2 =  0.363
    Difference (null H = exogenous): chi2(14)   =  19.37  Prob > chi2 =  0.151

********* xtabond or xtabond2 using a twostep procedure

xtabond  n  l(0/1).w   l(0/2).(k ys)  yr1978-yr1984  year, lags(2) twostep
* Note: The standard error from the twp step estimator is biased.  Thus, it is recommended to use the robust
*       to use the Windmeijer bias-corrected robust variance; below.

xtabond  n  l(0/1).w   l(0/2).(k ys)  yr1978-yr1984  year, lags(2) twostep robust
*xtabond  n  l(0/1).w   l(0/2).(k ys)  yr1978-yr1984 . year, lags(2) twostep vce(robust)

Arellano-Bond dynamic panel-data estimation  Number of obs      =       611
Group variable: id                           Number of groups   =       140
Time variable: year
                                             Obs per group:    min =         4
                                                               avg =  4.364286
                                                               max =         6

Number of instruments =     42               Wald chi2(16)      =    1104.72
                                             Prob > chi2        =     0.0000
Two-step results
```

48

```
                                (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6287089   .1934138     3.25   0.001     .2496248    1.007793
         L2. |  -.0651882   .0450501    -1.45   0.148    -.1534847    .0231084
             |
           w |
         --. |  -.5257597   .1546107    -3.40   0.001     -.828791   -.2227284
         L1. |   .3112899   .2030006     1.53   0.125     -.086584    .7091638
             |
           k |
         --. |   .2783619   .0728019     3.82   0.000     .1356728    .4210511
         L1. |   .0140994   .0924575     0.15   0.879     -.167114    .1953129
         L2. |  -.0402484   .0432745    -0.93   0.352    -.1250649    .0445681
             |
          ys |
         --. |   .5919243   .1730916     3.42   0.001      .252671    .9311776
         L1. |  -.5659863   .2611008    -2.17   0.030    -1.077734   -.0542381
         L2. |   .1005433   .1610987     0.62   0.533    -.2152043    .4162908
             |
      yr1979 |   .0151101   .0107102     1.41   0.158    -.0058816    .0361018
      yr1980 |    .030858   .0169166     1.82   0.068    -.0022979    .0640139
      yr1981 |  -.0096741   .0280446    -0.34   0.730    -.0646405    .0452922
      yr1982 |  -.0155376   .0209937    -0.74   0.459    -.0566845    .0256093
      yr1983 |   .0014798   .0148348     0.10   0.921    -.0275959    .0305554
        year |  -.0038946   .0061024    -0.64   0.523    -.0158551    .0080659
       _cons |   8.350532   12.49925     0.67   0.504    -16.14756    32.84862
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1979 D.yr1980 D.yr1981
                  D.yr1982 D.yr1983 D.year
Instruments for level equation
        Standard: _cons

**** Arellano-Bond test for zero autocorrelation in first-differenced errors
estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors
  +-----------------------+
  |Order |  z     Prob > z|
  |------+----------------|
  |   1  |-2.1255  0.0335 |
  |   2  |-.35166  0.7251 |
  +-----------------------+
   H0: no autocorrelation
```

49

```
**** Redo with xtabond2
xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys)  yr*, gmm(L.n) iv(w L.w  L(0/2).(k ys) yr*) twostep nolevel robus
* Note: xtbond2 provides the additional information on the exogeneity test on instrument subsets (difference

Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 41                      Obs per group: min =         4
Wald chi2(16) =   1104.72                                      avg =      4.36
Prob > chi2   =     0.000                                      max =         6
------------------------------------------------------------------------------
             |              Corrected
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .6287089   .1934138     3.25   0.001      .2496248    1.007793
         L2. |  -.0651882   .0450501    -1.45   0.148     -.1534847    .0231084
             |
           w |
         --. |  -.5257597   .1546107    -3.40   0.001     -.828791   -.2227284
         L1. |   .3112899   .2030006     1.53   0.125     -.086584    .7091638
             |
           k |
         --. |   .2783619   .0728019     3.82   0.000      .1356728    .4210511
         L1. |   .0140994   .0924575     0.15   0.879     -.167114     .1953129
         L2. |  -.0402484   .0432745    -0.93   0.352     -.1250649    .0445681
             |
          ys |
         --. |   .5919243   .1730916     3.42   0.001      .252671     .9311776
         L1. |  -.5659863   .2611008    -2.17   0.030     -1.077734   -.0542381
         L2. |   .1005433   .1610987     0.62   0.533     -.2152043    .4162908
             |
      yr1978 |   .0233675   .0366145     0.64   0.523     -.0483956    .0951306
      yr1979 |    .034583   .0335191     1.03   0.302     -.0311132    .1002793
      yr1980 |   .0464363    .030786     1.51   0.131     -.0139031    .1067757
      yr1981 |   .0020096   .0337365     0.06   0.952     -.0641127    .0681319
      yr1982 |  -.0077484   .0230417    -0.34   0.737     -.0529092    .0374124
      yr1983 |   .0053743   .0146532     0.37   0.714     -.0233453     .034094
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(w L.w k L.k L2.k ys L.ys L2.ys yr1976 yr1977 yr1978 yr1979 yr1980
    yr1981 yr1982 yr1983 yr1984)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/.).L.n
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -2.13  Pr > z =  0.034
Arellano-Bond test for AR(2) in first differences: z =  -0.35  Pr > z =  0.725
```

50

```
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(25)    =  67.59  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(25)    =  31.38  Prob > chi2 =  0.177
  (Robust, but can be weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:
  iv(w L.w k L.k L2.k ys L.ys L2.ys yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr198
> 3 yr1984)
    Hansen test excluding group:     chi2(11)    =  12.01  Prob > chi2 =  0.363
    Difference (null H = exogenous): chi2(14)    =  19.37  Prob > chi2 =  0.151


********* xtabond or xtabond2 when wages and capital are not strictly exogenous.

xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys)  yr*, gmm(L.(n w k)) iv(L(0/2).ys yr*) nolevel robust small
* Note: The level equation is not used; "nolevel"
* Note: gmm(L.(n w k)) implies that all of n, w and k are predetermined and they will be instrumented.

Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: id                              Number of obs      =       611
Time variable : year                            Number of groups   =       140
Number of instruments = 90                      Obs per group: min =         4
F(16, 140)    =     85.30                                      avg =      4.36
Prob > F      =     0.000                                      max =         6
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .8179867   .0859761     9.51   0.000     .6480073     .987966
         L2. |  -.1122756   .0502366    -2.23   0.027     -.211596   -.0129552
             |
           w |
         --. |  -.6816685   .1425813    -4.78   0.000    -.9635594   -.3997776
         L1. |   .6557083    .202368     3.24   0.001     .2556158    1.055801
             |
           k |
         --. |   .3525689   .1217997     2.89   0.004     .1117643    .5933735
         L1. |  -.1536626   .0862928    -1.78   0.077     -.324268    .0169428
         L2. |  -.0304529   .0321355    -0.95   0.345    -.0939866    .0330807
             |

**** Using the limited number of lags for GMM estimation (maxldep, maxlags)

xtabond  n  l(0/1).w   l(0/2).(k ys)  yr1978-yr1984  year, lags(2) twostep robust maxldep(3)
```

51

```
Number of instruments =     32                Wald chi2(16)        =    785.83
                                              Prob > chi2          =    0.0000
Two-step results
                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .4118672   .3457447     1.19   0.234    -.2657799    1.089514
         L2. |  -.0776316   .0484084    -1.60   0.109    -.1725103     .017247
             |
           w |
         --. |   -.439898   .1183368    -3.72   0.000    -.6718339    -.207962
         L1. |    .151073   .1757121     0.86   0.390    -.1933164    .4954624
     (omitted)

Instruments for differenced equation
        GMM-type: L(2/4).n
        Standard: D.w LD.w D.k LD.k L2D.k D.ys LD.ys L2D.ys D.yr1979 D.yr1980 D.yr1981
                  D.yr1982 D.yr1983 D.year
Instruments for level equation
        Standard: _cons

********* xtabond or xtabond2 when wages and capital are not strictly exogenous.
xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys)  yr*, gmm(L.(n w k)) iv(L(0/2).ys yr*) nolevel robust twostep
* Note: The level equation is not used; "nolevel"
* Note: gmm(L.(n w k)) implies that all of n, w and k are predetermined and they will be instrumented.

xtabond  n  l(0/2).(ys) yr*, pre(w, lag(1,.)) pre(k, lag(2,.)) lags(2) twostep robust
* close but not identical (not sure of this)


**** Using the limited number of lags for GMM estimation (maxldep, maxlags)

xtabond  n  l(0/1).w   l(0/2).(k ys)  yr1978-yr1984  year, lags(2) twostep robust maxldep(3)
xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys)  yr*, gmm(L.n, laglimit(1 3)) iv(w L.w  L(0/2).(k ys) yr*) twoste

xtabond  n  l(0/2).(ys) yr*, pre(w, lag(1,.)) pre(k, lag(2,.)) lags(2) twostep robust maxldep(3) maxlag(3)
xtabond2 n  L.n L2.n  w L.w  L(0/2).(k ys)  yr*, gmm(L.(n w k), laglimit(1 3)) iv(L(0/2).ys yr*) nolevel rob


********* system GMM using xtabond2

xtabond2 n L.n L(0/1).(w k) yr*, gmmstyle(L.(n w k)) ivstyle(yr*, equation(level)) robust small

     /* Instruments for first differences equation
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            L(1/.).(L.n L.w L.k)
        Instruments for levels equation
```

52

```
          Standard
            _cons
            yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            D.(L.n L.w L.k) */

xtabond2 n L.n L(0/1).(w k) yr*, gmmstyle(L.(n w k)) ivstyle(yr*) robust small

     /* Instruments for first differences equation
          Standard
            D.(yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984)
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            L(1/.).(L.n L.w L.k)
        Instruments for levels equation
          Standard
            _cons
            yr1976 yr1977 yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            D.(L.n L.w L.k)  */

     /*
       gmmstyle(varlist [, laglimits(# # ) collapse equation(fdiff | level | bothg) passthru split])
       ivstyle(varlist [, equation(fdiff | level | bothg) passthru mz ])

       For example, if w1 is predetermined and w2 endogenous, then instead of gmm(w1) gmm(w2, lag(2 .)),
       one could simply type  gmm(w1 L.w2).
     */

********* system GMM using xtdpd

xtdpd n L.n L(0/1).(w k) yr1978-yr1984, dgmm(w k n) lgmm(w k n) liv(yr1978-yr1984) vce(robust) two hascons

     /* Instruments for differenced equation
                GMM-type: L(2/.).w L(2/.).k L(2/.).n
        Instruments for level equation
                GMM-type: LD.w LD.k LD.n
                Standard: yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984 _cons
     */

xtabond2 n L.n L(0/1).(w k) yr1978-yr1984, gmm(L.(w k n)) iv(yr1978-yr1984, eq(level)) h(2) robust two small
     /* Instruments for first differences equation
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            L(1/.).(L.w L.k L.n)
        Instruments for levels equation
          Standard
            _cons
            yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984
          GMM-type (missing=0, separate instruments for each period unless collapsed)
            D.(L.w L.k L.n)
     */
```

[More Developments] mostly in dynamic panel model framework.

53

Question 1

If $w_{it-1}, w_{it-2}, \cdots$, can be IVs for Δw_{it},

Why not using $\Delta w_{it-1}, \Delta w_{it-2}, \cdots$?

($\Delta w_{it-1} = w_{it-1} - w_{it-2}, \cdots$)

check $E(\Delta w_{it-1} \Delta e_{it}) = 0$, also $\Delta w_{it-2}, \cdots$

But, Δw_{it-1} and Δw_{it} can be weakly correlated

Think about correlations:

$\text{Cov}(w_{it-1}, \Delta w_{it})$ ·· suggested IVs

$\text{Cov}(\Delta w_{it-1}, \Delta w_{it})$ ·· can be weak, and it can lead to large std. errors of estimators.

But, can we use Δw_{it} as IV for w_{it}? (swap them!!)

Question 2

Do we need to use the FD method always?

How about using the level equation (1), and use Δw_{it} as IVs? (swap!!) ⇒ Use this especially when ρ is close to 1. (persistent!)

Estimation: $y_{it} = \alpha_i + \rho y_{it-1} + X_{it}\gamma + w_{it}\delta + R_{it}c + e_{it}$ (1)

IVs ⇒ y_{it-1}: $\Delta y_{it-1}, \Delta y_{it-2}, \vdots$; w_{it}: $\Delta w_{it}, \Delta w_{it-1}, \Delta w_{it-2}, \vdots$; R_{it}: $\Delta R_{it-1}, \Delta R_{it-2}, \vdots$

$$\begin{cases} E(\Delta w_{it} e_{it}) = E(w_{it} e_{it}) + E(w_{it-1} e_{it}) = 0 \\ E(\Delta R_{it-1} e_{it}) = E(R_{it-1} e_{it}) + E(R_{it-1} e_{it}) = 0 \end{cases}$$

But, how can we control for α_i without using FEs?

It can be done.

⇒ "Level GMM" Estimator & System GMM Estimator

(Call the FD based GMM as "difference GMM")
(initial Arellano & Bond)

(Forward)
Orthogonal transformation for level GMM estimator

Arellano & Bover (1995)

$$w_{it}^* = c_{it}\left[w_{it} - \frac{1}{T-t}\sum_{s>t} w_{is}\right]$$

$\sum_{s>t} w_{is}$: $w_{it+1} + w_{it+2} + \cdots + w_{iT}$ (forward values)

where $c_{it} = \sqrt{\dfrac{T-t}{T-t+1}}$

.. forward orthogonal deviation (alternative to FE)

point : $E(v_{it}^* e_{it}) = 0$

$\Rightarrow$ Instead of using FE method, we transform the data using the forward orthogonal deviation.

point : How about α_i ?

$\Rightarrow$ We assume $E(\Delta w_{it}\, \alpha_i) = 0$ for all i & t

$\Rightarrow$ $E(w_{it}\, \alpha_i)$ is time-invariant.

$\Rightarrow$ Δw_{it-1} is a valid IV for w_{it}

$$\left[\begin{array}{l} E[\Delta w_{it-1} e_{it}) = E[\Delta w_{it}\alpha_i] + E[w_{it-1} v_{it}] \\ \qquad\qquad\qquad - E[w_{it-2} v_{it}] \\ \qquad\qquad\qquad = 0 + 0 - 0 \\ \text{where } e_{it} = \alpha_i + v_{it} \end{array}\right]$$

$\Rightarrow$ use Δw_{it-1} as IV for w_{it}.

Note The above assumption is related to the assumption on the initial value of w_{it}..

$w_{it} = \alpha w_{it-1} + \alpha_i + v_{it}$

"Fast growing firms are not systematically closer or farther from their steady states than slower growing ones" $\Rightarrow$

$\rightarrow$ converges to $\alpha_i/(1-\alpha)$.

If $E(w_{it}\alpha_i)$ is time-invariant,

$E[\alpha_i(w_{it} - \alpha_i/(1-\alpha))] = 0$

(some story to explain this : read the paper by A.B. (1995))

System GMM

Combine : difference GMM IVs

+ level GMM IVs

$$\Delta y_{it} = \rho \underbrace{\Delta y_{it-1}} + \Delta x_{it}\gamma + \cdots + \Delta e_{it} \qquad \underline{\text{difference}}$$

$$\Downarrow \; IV = y_{it-2} \; ; \; E(y_{it-2}\Delta e_{it}) = 0$$

$$y_{it} = \rho \underbrace{y_{it-1}} + x_{it}\gamma + \cdots + e_{it} \qquad \underline{\text{level}}$$

$$\Downarrow \; IV = \Delta y_{it-1}$$

$$; \; E(\Delta y_{it-1} e_{it}) = 0$$

<u>STATA</u> xtdpdsys n L(0/2).(w k) yr* , robust

.... dep.

i) IV for differenced equation : L(2/.).n

$IV = (y_{it-2}, \cdots, y_{i1})$ for Δy_{it-1}

ii) IV for level equation : LD.n

$IV = \Delta y_{it-1}$ for y_{it-1}

<u>STATA</u> xtabond2 .... , ~~nolevel~~

Delete this option.

(default is system GMM)

<u>STATA</u> xtdpd .. (more general than xtdpdsys or xtband)

- dgmmiv .. IVs for dif. eq
- lgmmiv ...IVs for level eq
- iv ... strictly exog in both eq
- div ... " in diff eq
- liv .. " in level eq

How to combine them? (for system GMM)

i) Stack the transformed (FD or Forward orth. deviation) variable and the untransformed variables together

- $Y_i = \begin{bmatrix} y_i^* \\ y_i \end{bmatrix}$ where y_i^* is the transformed data

$y_i^* \Rightarrow y_{it} - y_{it-1}$ (FD)

$\left[y_{it} - \left(\sum_{s>t}^{T} y_{is}\right)/T-t\right] c_{it}$

(F.O.D) Forward orthog. deviation

(y_i = raw data)

- $X_i = \begin{bmatrix} x_i^* \\ x_i \end{bmatrix}$

where x_i^* = transformed data (FD or FOD)

x_i = raw data (indep. variables)

- Z_i (= IV matrix)

$$= \begin{bmatrix} Zd_i & 0 & D_i & 0 & I_i^d \\ 0 & ZL_i & 0 & L_i & I_i^L \end{bmatrix}$$

where
- Zd_i = GMM type IVs from dgmmiv
- ZL_i = " lgmmiv
- D_i = IVs (strictly exog) from div
- L_i = " from liv
- I_i^d = IVs for diff. eq from iv
- I_i^L = " level eq "

ii) Then follow a similar procedure as in Arellano & Bond GMM estimator. (p5 or p.19 of Lecture 6)

<u>Note</u> System GMM can lead to bias, since it uses <u>many</u> IVs. Using the limited # of IVs (lags) would be preferred!

57

```
* panel_2_and_dynamic.do

capture log close
log using panel_2_and_dynamic.log, replace

* Data and code: from
* "Microeconometrics using Stata", 2008 by A. Colin Cameron and Pravin K. Trivedi
* Chapter 9
* 9.3: HAUSMAN TAYLOR ESTIMATOR
* 9.4: ARELLANO-BOND ESTIMATOR

* To run you need files
*   panel_2_wage.dta
* in your directory
* No Stata user-written commands are used

********** SETUP **********

set more off
clear all

set memory 30m
set linesize 90

********** DATA DESCRIPTION **********

* mus08psidextract.dta -> panel_2_wage.dta
* PSID. Same as Stata website file psidextract.dta
* Data due to Baltagi and Khanti-Akom (1990)
* This is corrected version of data in Cornwell and Rupert (1988).
* 595 individuals for years 1976-82

use panel_2_wage.dta, clear

******* HAUSMAN-TAYLOR ESTIMATOR

* Hausman-Taylor example of Baltagi and Khanti-Akom (1990)

 xthtaylor lwage occ south smsa ind fem blk  exp exp2 wks ms union ed,  ///
  endog(exp exp2 wks ms union ed)

// Hausman-Taylor with panel bootstrap SEs or with jackknife
* xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, ///
*  endog(exp exp2 wks ms union ed) vce(boot, reps(400) nodots seed(10101))
* xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed, ///
*  endog(exp exp2 wks ms union ed) vce(jackknife)

******* ARELLANO-BOND ESTIMATOR

* 2SLS or one-step GMM for a pure time-series AR(2) panel model
```

58

```
xtabond lwage, lags(2) vce(robust)

* Optimal or two-step GMM for a pure time-series AR(2) panel model

xtabond lwage, lags(2) twostep vce(robust)

* Reduce the number of instruments for a pure time-series AR(2) panel model

xtabond lwage, lags(2) vce(robust) maxldep(1)

* Optimal or two-step GMM for a dynamic panel model

xtabond lwage occ south smsa ind, lags(2) maxldep(3)     ///
 pre(wks,lag(1,2))                                        ///
 endogenous(ms,lag(0,2))                                  ///
 endogenous(union,lag(0,2))                               ///
 twostep vce(robust) artests(3)

* Test whether error is serially correlated

estat abond

******* ARELLANO-BOVER SYSTEM ESTIMATOR

* Arellano/Bover or Blundell/Bond for a dynamic panel model

xtdpdsys lwage occ south smsa ind, lags(2) maxldep(3)    ///
 pre(wks,lag(1,2))                                        ///
 endogenous(ms,lag(0,2))                                  ///
 endogenous(union,lag(0,2))                               ///
 twostep vce(robust) artests(3)

estat abond

* Test of overidentifying restrictions (with no vce(robust))

quietly xtabond lwage occ south smsa ind, lags(2) maxldep(3) ///
 pre(wks,lag(1,2))                                        ///
 endogenous(ms,lag(0,2))                                  ///
 endogenous(union,lag(0,2))                               ///
 twostep artests(3)
estat sargan

******* xtdpd

* Use of xtdpd to exactly reproduce the previous xtdpdsys command

xtdpd L(0/2).lwage  L(0/1).wks  occ south smsa ind ms union, ///
```

59

```
  div(occ south smsa ind)                                     ///
  dgmmiv(lwage, lagrange(2 4))                                ///
  dgmmiv(ms union, lagrange(2 3))                             ///
  dgmmiv(L.wks, lagrange(1 2))                                ///
  lgmmiv(lwage wks ms union)                                  ///
  twostep vce(robust) artests(3)

* Previous command if model error is MA(1)

 xtdpd L(0/2).lwage  L(0/1).wks  occ south smsa ind ms union, ///
  div(occ south smsa ind)                                     ///
  dgmmiv(lwage, lagrange(3 4))                                ///   ***
  dgmmiv(ms union, lagrange(2 3))                             ///
  dgmmiv(L.wks, lagrange(1 2))                                ///
  lgmmiv(lwage wks ms union)                                  ///
  twostep vce(robust) artests(3)

log close
```

Other panel data models

(1) paired (matched) sample (if x_{it} are strictly exog.)

$$\begin{cases} y_{i1} = x_{i1}\beta + f_i + u_{i1} \\ y_{i2} = x_{2i}\beta + f_i + u_{2i} \end{cases}$$

f_i = unobserved family effect.

- Do usual RE as in $T=2$.
- Do FD over siblings as in $T=2$
  = FE as in $T=2$.

  : Ashenfelter & Krueger (1994)

(2) Cluster samples

$$y_{is} = x_{is}\beta + c_i + u_{is}$$

c_i : cluster (unobserved effects)

- Do FE if c_i is correlated with x_{is}.
- Can do RE.

Note within-group correlation can be additionally considered : XTGEE in Stata.

Exercise Ex 11.11 JTRAIN1.DTA. (Do FD & FE and compare)

Note First difference

```
gen cclscrap = clscrap - clscrap[_n-1] if d89
```

:

Note FE should give the same result as in FD.

```
xtreg clscrap d89 cgrant cgrant_1, fe
```

: "The results may give wrong signs, though."

Exercise Ex 11.15 JTRAIN1.DTA.

eg 2) Levitt's prison population effects on crime rate "prison.dta"

$$\Delta \log crime_{it} = \Delta X_{it}\beta + \Delta \log(prism_{it})\gamma + \underbrace{\gamma_t}_{\text{time dummies.}} + \Delta u_{it}$$

; α_i was differenced away.

IV for $\Delta \log(prism_{it})$ = two dummy variables

$\left\{ \begin{array}{l} D_{1it} = 1, \text{ a final decision was reached on overcrowding litigation in the current year} \\ D_{2it} = 1, \text{ if a final decision was reached in the previous two years.} \end{array} \right.$

✗ Modified Panel Data Models

Models with Individual-Specific Slopes (No dynamic or feedback effects)

(1) $y_{it} = \alpha_i + \gamma_t + x_{it}\beta + u_{it}$

α_i: individual dummies; γ_t: time dummies $(D_{1t}, D_{2t}, \ldots, D_{T-1,t})$: T-1 dummies

This is a usual time FE model

(2) $y_{it} = \alpha_i + g_i t + \gamma_t + x_{it}\beta + u_{it}$

.. Random trend model, individual specific trend

eg. state-specific trend, in addition to state-specific dummies (α_i)

$$\Rightarrow \Delta y_{it} = (\cancel{\alpha_i} - \cancel{\alpha_i}) + g_i + \Delta X_{it}\beta + \Delta u_{it}$$

.. FE in differences!

eg) Friedberg (1998, AER) ($state_i * time_t$)

: The results are changed.

(3) Trend in union participation (interaction trend) ✱

$$y_{it} = \alpha_i + \gamma_t + g_i t + X_{it}\beta + \delta_1 Union_{it} + \delta_2 (Union_{it} \cdot t) + u_{it}$$

$$\Delta y_{it} = \xi_t + g_i + \Delta X_{it}\beta + \delta_1 \Delta Union_{it} + \delta_2 \Delta(Union_{it} \cdot t) + \Delta u_{it}$$

FD with FE $\overset{\text{same}}{\Leftrightarrow}$ FD twice (SD)

(4) Time constant unobserved heterogeneity

$$y_{it} = x_{it}\beta + \alpha_{1i} + \alpha_{2i} * prog_{it} + u_{it}$$

where $prog_{it} = \begin{cases} 1 & \text{participated} \\ 0 & \text{not.} \end{cases}$

eg) Lemieux (JLE, 1998)
Unobserved heterogeneity is rewarded differently in the union and non-union sectors.

(No Feedback is assumed in the next models)

Hausman & Taylor Models

Suppose that we have

i) time-invariant variables → FE cannot be used.

ii) some regressors are correlated with α_i → RE cannot be used

Hausman & Taylor method can be used in this case.

Let

$$y_{it} = X_{it}\beta + Z_i\delta + \mu_i + \varepsilon_{it}$$

where $X_{it} = \begin{cases} X_{1it} & : \text{exog. time variant} \\ X_{2it} & : \text{correlated with } \mu_i \text{ but not with } \varepsilon_{it} \end{cases}$

$Z_i = \begin{cases} Z_{1i} & : \text{exog. time invariant.} \\ Z_{2i} & : \text{correlated with } \mu_i \text{ but not with } \varepsilon_{it} \end{cases}$

Thus $E(X_{1it}'\mu_i) = 0$, $E(Z_{1i}'\mu_i) = 0$

these are exogenous, thus can be used as IVs.

Fama & MacBeth ((1973, JPE)

Method (cf. Asset pricing model)

R_{it} = rate of return of each stock i at period t

R^M_t = market rate of return

1st Regress and estimate β_i by time series on each stock i

$$R_{it} - r_t = \alpha_i + \beta_i (R^M_t - r_t) + \varepsilon_{it}$$

using T observations $\Rightarrow \hat{\beta}_i$, $i = 1, \dots, N$

2nd Run cross-section regression

$$\bar{R}_i = \alpha_0 + \alpha_1 \hat{\beta}_i + u_i$$

using N observations. ($\bar{R}_i$ = average returns of each stock i

CAPM implies $\alpha_0 = \bar{r}$, $\alpha_1 = \bar{R}_M - \bar{r}$

Additional test

$$\bar{R}_i = \alpha_0 + \alpha_1 \hat{\beta}_i + \alpha_2 \underbrace{\hat{\beta}_i^2}_{\text{nonlinearity}} + \alpha_3 \underbrace{\hat{\sigma}^2_{\varepsilon i}}_{\text{firm-specific risk}} + w_i$$

CAPM implies $\alpha_2 = \alpha_3 = 0$

Two issues

i) Measurement error $\hat{\beta}_i = \beta_i + (\text{something}) * \varepsilon_i$

ii) Heteroskedasticity

Review

(Panel data Models II)

1. "Feedback" effects in panel data models.

   a) what are they? Define the feedback effects.

   b) what are the problems of the usual FE, RE & FD estimators when feed back effects are present?

   c) what are the proposed solutions?

2. Hausman & Taylor models

   a) When are the Hausman-Taylor models useful?

   b) What is the required condition for identification?

   c) Discuss briefly the estimation procedure.

3. Dynamic Panel models

$$y_{it} = \alpha_i + \beta y_{it-1} + X_{it}\gamma + W_{it}\delta + R_{it}C + e_{it} \quad \begin{pmatrix} i=1,\cdots,N \\ t=1,\cdots,T \end{pmatrix}$$

where X_{it} is strictly exogenous

W_{it} is predetermined

R_{it} is endogenous (IVs = Z_{it} for R_{it})

One considers to estimate the above model using the first differenced data.

   a) What are the IVs for Δy_{it-1} that Anderson-Hsiao considered?

b) What are the IVs for Δy_{it-1} that Arellano-Bond considered ?

c) Find ALL possible IVs for each of

i) Δy_{it-1}

ii) ΔW_{it}

iii) ΔR_{it}

d) Define each of strict exogenous, predetermined and endogenous regressors.

4. Alternatively, one considers to keep the level equation in 3, without taking the first difference of the data to estimate the dynamic panel data model.

a) What are the IVs that Arellano-Bover suggested ?

b) Find ALL possible IVs for

i) y_{it-1}

ii) W_{it}

iii) R_{it}

c) What is the system GMM estimator ? How is it defined ?

d) Explain the assumption on d_i when the level GMM is employed.

5. Explain

a) Why testing for autocorrelation in the Arellano-Bond (A-B) estimation is important.

b) Why autocorrelation at lag 1 tends to exist but not at lag 2 if the model is correctly specified.

c) How you can modify the A-B model if autocorrelation is present at lag 2.

6. Dynamic panel data models

a) How can you test for the need to consider a dynamic model, compared with a static model which you may consider initially?

b) Under what conditions, the usual FE, RE & FD estimators would be biased? When bias would disappear?

c) Why the bias problem, if it exists, would be more serious in FD estimators?

d) Why using more IVs (moment conditions) would be good? Why bad? Depending on what?