面板数据的计量经济分析

王群勇 (教授, 博士生导师)

南开大学　数量经济研究所

Contents

Balanced panel

- Some methods require balanced panel.
  (1) Amemiya–MaCurdy estimator is available for balanced panels
  (2) Fixed effect panel threshold model

example

- data: "bofdi.dta"
- step 1: view the panel pattern

```
. use "bofdi", clear
. local varlist "lnofdi lngdp lngdpca lngdpch lndist lnpd"
. xtdes

      id:  1, 2, ..., 64                                   n =         64
    year:  2003, 2004, ..., 2014                           T =         12
           Delta(year) = 1 unit
           Span(year)  = 12 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%      25%       50%       75%     95%     max
                        12      12       12        12        12      12      12

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+--------------
       64    100.00  100.00 |  111111111111
 ---------------------------+--------------
       64    100.00         |  XXXXXXXXXXXX
```

example

- step 2: drop observation with missings

```
. gen mi = mi(lnofdi,lngdp,lngdpca,lngdpch,lndist,lnpd)
. drop if mi
(406 observations deleted)
. xtdes

      id:  2, 3, ..., 62                                    n =         46
    year:  2003, 2004, ..., 2013                            T =         11
           Delta(year) = 1 unit
           Span(year)  = 11 periods
           (id*year uniquely identifies each observation)
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       7         9        10      10      11

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-------------
       21     45.65   45.65 |  1111111111.
        3      6.52   52.17 |  .1.1111111.
        3      6.52   58.70 |  11.1111111.
        2      4.35   63.04 |  .........1.
        2      4.35   67.39 |  ...1111111.
        2      4.35   71.74 |  ..11111111.
        1      2.17   73.91 |  .....11111.
        1      2.17   76.09 |  ....1...11.
        1      2.17   78.26 |  ....1.1111.
       10     21.74  100.00 | (other patterns)
 ---------------------------+-------------
       46    100.00         |  XXXXXXXXXXX
```

example

- step 3: fill the gap for the specified period

```
. local tmin = 2005
. local tmax = 2012
. keep if inrange(year, `tmin', `tmax')
(64 observations deleted)
. tsfill, full
. xtdes
     id:  2, 3, ..., 62                                        n =          46
   year:  2005, 2006, ..., 2012                                 T =           8
          Delta(year) = 1 unit
          Span(year)  = 8 periods
          (id*year uniquely identifies each observation)
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       7         8         8       8       8

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+----------
       26     56.52   56.52 |  11111111
        8     17.39   73.91 |  .1111111
        2      4.35   78.26 |  .......1
        1      2.17   80.43 |  ....1111
        1      2.17   82.61 |  ...11111
        1      2.17   84.78 |  ..1...11
        1      2.17   86.96 |  ..1.1...
        1      2.17   89.13 |  ..1.1111
        1      2.17   91.30 |  .1...1..
        4      8.70  100.00 | (other patterns)
 ---------------------------+----------
       46    100.00         |  XXXXXXXX
```

example

- step 4: select the balanced panel

```
. replace mi = mi(lnofdi,lngdp,lngdpca,lngdpch,lndist,lnpd)
(62 real changes made)
. by id: egen nmi = total(mi)
. keep if nmi==0
(160 observations deleted)
. xtdes

      id:  2, 3, ..., 62                                    n =         26
    year:  2005, 2006, ..., 2012                            T =          8
           Delta(year) = 1 unit
           Span(year)  = 8 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         8       8       8         8         8       8       8

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+----------
       26    100.00  100.00 |  11111111
 ---------------------------+----------
       26    100.00         |  XXXXXXXX
```

example

- ado program

```
program mybalan
syntax varlist, tlist(numlist)
local vc : subinstr local varlist " " ",", all
tempvar mi nmi
qui gen `mi' = mi(`vc')
qui drop if `mi'

qui xtset
local pvar = r(panelvar)
local tvar = r(timevar)
qui numlist "`tlist'"
local nst = r(numlist)
local nst: subinstr local nst " " ",", all
qui keep if inlist(`tvar', `nst')

tsfill, full
replace `mi' = mi(`vc')
by `pvar': egen `nmi' = total(`mi')
keep if `nmi'==0
xtdes
end
```

example

- example

```
. use "bofdi.dta", clear
. mybalan lnofdi lngdp lngdpca lnpd, tlist(2005/2012)
```

Contents

panel

- error component model

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

  c_i: individual heterogeneity.
  $v_i = c_i + u_{it}$: combined error.

- How to estimate β?
  1. $E(\mathrm{x}'c) = 0$, random effect
  2. $E(\mathrm{x}'c) \neq 0$, fixed effect.

random effect model

- Pooled OLS.
- GLS.

$$
\begin{aligned}
Var(v_{it}) &= \sigma_c^2 + \sigma_u^2 \\
Cov(v_{it} v_{is}) &= E[(c_i + u_{it})(c_i + u_{is})] = \sigma_c^2
\end{aligned}
$$

$$\rho_{ts} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_u^2}$$

GLS transformation: $\tilde{y}_{it} = y_{it} - \theta_i \bar{y}_i$ where

$$\theta_i = 1 - \sqrt{\frac{\sigma_u^2}{T_i \sigma_c^2 + \sigma_u^2}}.$$

fixed effect model

- difference estimation

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

Assumption (strict exogeneity):

$$E[(\mathrm{x}_{it} - \mathrm{x}_{it-1})'(u_{it} - u_{it-1})] \neq 0.$$

- within group estimation (fixed effect estimation)

$$\bar{y}_i = \mu + \bar{\mathrm{x}}_i\beta + c_i + \bar{u}_i$$

$$y_{it} - \bar{y}_i = (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\beta + u_{it} - \bar{u}_i$$

Assumption (strict exogeneity):

$$E[(\mathrm{x}_{it} - \bar{\mathrm{x}}_i)'(u_{it} - \bar{u}_i)] \neq 0.$$

fixed effect model

- If $T = 2$, FE estimator is equivalent to FD.

$$y_{i2} - \frac{y_{i1} + y_{i2}}{2} = (\mathrm{x}_{i2} - \frac{\mathrm{x}_{i1} + \mathrm{x}_{i2}}{2})\beta + u_{i2} - \frac{u_{i1} + u_{i2}}{2}.$$

  For $T > 2$, FE estimator is more efficient.

- Hausman test:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \sim \chi^2(K).$$

Syntax

- command

```
. xtreg dep varlist, options
. xttest0 * random effect test
. hausman consistent efficient
```

options include `fe`, `re`, `be`, `mle`, `pa`.

example

- example (abdata.dta)

```
xtreg n L(0/2).(w k) yr1980-yr1984 year, fe
est store fe
xtreg n L(0/2).(w k) yr1980-yr1984 year, re
est store re
xttest0
hausman fe re
hausman fe re, sigmamore
```

Contents

panel

- panel:
$$y_{it} = \mu + \mathrm{x}_{it}\beta + c_i + u_{it}$$
two endogeneity sources: c_i, u_{it}.
- If $E(x'_{it}u_{it}) \neq 0$,$E(x'_{it}c_i) = 0$: random effect iv.
- If $E(x'_{it}u_{it}) \neq 0$,$E(x'_{it}c_i) \neq 0$: fixed effect iv or first difference iv.

dynamic panel

- dynamic model:

$$y_{it} = \mu + \rho y_{i,t-1} + \mathrm{x}_{it}\beta + c_i + u_{it}$$

- difference estimation

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

  Problem:

$$E[(y_{it-1} - y_{it-2})(u_{it} - u_{it-1})] \neq 0.$$

- IV estimation: use $y_{it-2}, y_{it-3}, \ldots$ as IVs.

dynamic panel

- Arrelano and Bond (1991): use GMM-type IVs.

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

for $t = 3$, $y_{i2} - y_{i1}$, $u_{i3} - u_{i2}$. IV: y_{i1}
for $t = 4$, $y_{i3} - y_{i2}$, $u_{i4} - u_{i3}$. IV: y_{i2}, y_{i1}
for $t = 5$, $y_{i4} - y_{i3}$, $u_{i5} - u_{i4}$. IV: y_{i3}, y_{i2}, y_{i1}
......

- number of IVs: $(T-2)(T-1)/2 + K + 1$.
- Problem: GMM estimators with too many overidentifying restrictions may perform poorly in small samples (Kiviet, 1995). (1) multicollinearity among IVs. (2) Weak IVs.
- Solution: choose IVs with fixed lags.

- system GMM (Arellano and Bover, 1995; Blundell and Bond, 1998): IVs include lagged levels as well as lagged differences.
- Differenced residuals should not exhibit significant AR(2) behavior.

$$Var(\Delta u_{it}) = E[(u_{it} - u_{it-1})^2] = 2\sigma_u^2$$

$$\begin{aligned} E[(u_{it} - u_{it-1})(u_{it-1} - u_{it-2})] &= E[-u_{it-1}^2] = -\sigma_u^2 \\ E[(u_{it} - u_{it-1})(u_{it-2} - u_{it-3})] &= 0 \end{aligned}$$

So,

$$\rho_1 = -0.5, \quad \rho_k = 0 (k \geq 2)$$

Syntax

- command for Arrelano and Bond GMM estimation

```
. xtabond dep varlist, options
. xtdpdsys dep varlist,options
```

options include:

- `lags`(*num*): maximum lag of y_{it} as independent (1 by default).
- `maxldep`(*num*): maximum number of lags as IVs (all lags by default).
- `twostep`: two-step estimation.
- `pre`(*varlist*, ...)
- `endog`(*varlist*, ...)

example

- example (abdata.dta)

```
xtabond  n L(0/2).(w k) yr1980-yr1984 year, vce(robust)
xtdpdsys n L(0/2).(w k) yr1980-yr1984 year, vce(robust)
estat abond, artest(3)
estat sargan
```

dynamic panel

- more lags

$$\Delta y_{it} = \rho_1 \Delta y_{i,t-1} + \rho_2 \Delta y_{i,t-2} + \Delta \mathrm{x}_{it}\beta + \Delta u_{it}$$

for $t = 3$, $y_{i2} - y_{i1}$, $u_{i3} - u_{i2}$. IV: y_{i1}
for $t = 4$, $y_{i3} - y_{i2}$, $y_{i2} - y_{i1}$, $u_{i4} - u_{i3}$. IV: y_{i2}, y_{i1}
for $t = 5$, $y_{i4} - y_{i3}$, $y_{i4} - y_{i3}$, $u_{i5} - u_{i4}$. IV: y_{i3}, y_{i2}, y_{i1}
......

dynamic panel

- predetermined explanatory variable: $E(\mathrm{x}_{i,t+1}u_{it}) \neq 0.$

$$\Delta y_{it} = \rho_1 \Delta y_{i,t-1} + \Delta \mathrm{x}_{it}\beta + \Delta u_{it}$$

for $t = 3$, $y_{i2} - y_{i1}$, $x_{i3} - x_{i2}$, $u_{i3} - u_{i2}$. IV: y_{i1}, x_{i2}, x_{i1}
for $t = 4$, $y_{i3} - y_{i2}$, $x_{i4} - x_{i3}$, $u_{i4} - u_{i3}$. IV: $y_{i2}, y_{i1}, x_{i3}, ..., x_{i1}$
for $t = 5$, $y_{i4} - y_{i3}$, $x_{i5} - x_{i4}$, $u_{i5} - u_{i4}$. IV: $y_{i3}, y_{i2}, y_{i1}, x_{i4}, ..., x_{i1}$
......

dynamic panel

- endogenous explanatory variable: $E(\mathrm{x}_{it}u_{it}) \neq 0$.

$$\Delta y_{it} = \rho_1 \Delta y_{i,t-1} + \Delta \mathrm{x}_{it}\beta + \Delta u_{it}$$

for $t = 3$, $y_{i2} - y_{i1}$, $x_{i3} - x_{i2}$, $u_{i3} - u_{i2}$. IV: y_{i1}, x_{i1}
for $t = 4$, $y_{i3} - y_{i2}$, $x_{i4} - x_{i3}$, $u_{i4} - u_{i3}$. IV: $y_{i2}, y_{i1}, x_{i2}, x_{i1}$
for $t = 5$, $y_{i4} - y_{i3}$, $x_{i5} - x_{i4}$, $u_{i5} - u_{i4}$. IV: $y_{i3}, y_{i2}, y_{i1}, x_{i3}, x_{i2}, x_{i1}$
......