# Empirical Panel Data: Lecture 11

INSTRUCTOR: CHAOYI CHEN
NJE & MNB

May 16, 2023

# Topic 4: An empirical example for the dynamic panel regression

- Recall, we consider the following augmented dynamic panel AR(1) panel as an example before

$$\Delta G_{it} = \beta_0 + \gamma \Delta G_{it-1} + \beta_1 \Delta C_{it} + \beta_2 \Delta K_{it} + \beta_3 \Delta P_{it} + \alpha_i + \varepsilon_{it}.$$

- $\Delta G_{it}$ is the economic growth rate
- $\Delta C_{it}$ is the consumption growth rate
- $\Delta K_{it}$ is the investment growth rate
- $\Delta P_{it}$ is the population growth rate

# Topic 4: Stata command for first-differenced GMM and system GMM

- To perform the Arellano and Bond GMM estimation,
  `xtabond2 depvar varlist [if] [in] [weight], [, level() twostep robust cluster nonconstant small nolevel]` *ivstyle() gmmstyle()*

- To perform the system GMM estimation,
  `xtabond2 depvar varlist [if] [in] [weight], [, level() twostep robust cluster nonconstant small` ~~`nolevel`~~`]` *ivstyle() gmmstyle()*

# A Stata example: Arellano and Bond estimator

- xtabond2 lrgdpnagrowth L.lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, gmm(L.lrgdpnagrowth) nolevel robust small

```
Dynamic panel-data estimation, one-step difference GMM
------------------------------------------------------------------------------
Group variable: country1                        Number of obs      =      8870
Time variable : year                            Number of groups   =       180
Number of instruments = 1953                    Obs per group: min =        22
F(0, 180)     =          .                                     avg =     49.28
Prob > F      =          .                                     max =        62
------------------------------------------------------------------------------
             |               Robust
lrgdpna_growth|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lrgdpna_growth|
         L1. |   .1453177   .0252345     5.76   0.000     .0955243    .1951111
             |
 lccon_growth |   .3965805   .0488324     8.12   0.000     .3002229    .4929382
   lck_growth |   .0031694   .0358864     0.09   0.930    -.0676428    .0739817
  lpop_growth |   .6539932    .213688     3.06   0.003     .2323375    1.075649
------------------------------------------------------------------------------
Instruments for first differences equation
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/64).L.lrgdpna_growth
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -6.53  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.17  Pr > z =  0.867
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(1949) =2690.42  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
```

# A Stata example: Blundell and Bond system GMM estimator

- `xtabond2 lrgdpnagrowth L.lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, gmmstyle(L.lrgdpnagrowth) robust small`

```
Dynamic panel-data estimation, one-step system GMM
------------------------------------------------------------------------------
Group variable: country1                        Number of obs      =      9050
Time variable : year                            Number of groups   =       180
Number of instruments = 2016                    Obs per group: min =        23
F(4, 179)      =   1004.17                                     avg =     50.28
Prob > F       =     0.000                                     max =        63
------------------------------------------------------------------------------
               |              Robust
lrgdpna_growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
lrgdpna_growth |
           L1. |    .155933   .0251676     6.20   0.000     .1062696    .2055965
               |
  lccon_growth |   .4244219   .0459817     9.23   0.000      .333686    .5151578
    lck_growth |    .017377   .0365581     0.48   0.635    -.0547632    .0895172
   lpop_growth |   .4232537   .1152593     3.67   0.000     .1958119    .6506956
         _cons |   .5818639   .2837831     2.05   0.042     .0218732    1.141855
------------------------------------------------------------------------------
Instruments for first differences equation
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(1/64).L.lrgdpna_growth
Instruments for levels equation
  Standard
    cons
```

# Topic 5: Panel unit root test

- Considering the following AR(1) model

$$y_t = \gamma y_{t-1} + \varepsilon_t, \ \varepsilon \sim WN(0, \sigma^2).$$

- If $|\gamma| = 1$, $y_t$ is a unit root. One can apply ADF (Augmented Dicky-Fuller) test to test the unit root.
- However, for the dynamic panel AR(1) model

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}.$$

- We wish to test if $\gamma = 1$.
- However, it is well established that the ADF test shows a significantly low power.
- Need to develop panel unit root test!

# Topic 5: First-generation panel unit root test

- Earlier studies on panel unit root tests assume cross-sectional **independence**. That is $\varepsilon_{it}$ is independent across $i$.
- Among the all first-generation panel Unit root tests in the literature, we will focus on following two testings:
  1. LLC test (Levin *et al.* 2002)
  2. IPS test (Im *et al.* 2003)
- Consider the hetetrogeneous dynamic panel model

$$y_{it} = \gamma_i y_{it-1} + \alpha_i + \varepsilon_{it}.$$

- Subtracting both sides by $y_{it-1}$ yields

$$\Delta y_{it} = \delta_i y_{it-1} + \alpha_i + \varepsilon_{it},\ \delta_i = (\gamma - 1).$$

- LLC and IPS tests examine $H_0 : \delta_i = 0$ for all $i$ against $H_a : \delta_i \leq 0$.
- However, they differ in whether the alternative holds for all $i$.

# Topic 5: LLC test

- LLC focus on the following generalized equation:

$$\Delta y_{it} = \delta y_{it-1} + \alpha_i + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \theta_t + \beta_i t + \varepsilon_{it}$$

- Remark 1: LLC assumes all $\delta_i = \delta$ for all $i$.
- Remark 2: LLC assumes the errors are independently distributed across panels.
- Remark 3: Similar to ADF test, LLC propose different specifications, allowing additional lags of dependent variables and the time trend.

- LLC propose the following hypothesis testing:

$$H_0 : \ \delta = 0$$
$$H_a : \ \delta < 0$$

# Topic 5: LLC test in practice

- LLC propose using a two-steps to contruct the test statistic (w.l.o.g. we assume no time trend and no time fixed effects):
  1. Step 1: conduct the first-stage regression and generate orthogalized residuals

     $$1).\ \Delta y_{it} = \alpha_i + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + e_{it},\ \ 2).\ y_{it} = \gamma_i + \sum_{j=1}^{p_i} \phi_{ij} \Delta y_{i,t-j} + v_{it}.$$

     Then standardize the residuals: $\tilde{e}_{it} = \frac{\widehat{e}_{it}}{\widehat{\sigma}_{\varepsilon_i}},\ \ \tilde{v}_{it} = \frac{\widehat{v}_{it}}{\widehat{\sigma}_{\varepsilon_i}}.$
  2. Step 2: Using the normalized residuals to run the pooled ADF regression

     $$\tilde{e}_{it} = \beta \tilde{v}_{it-1} + w_{it}.$$
- Therefore, testing the null hypothesis is the same as testing if $\beta = 0$.
- However, the studentized coefficient $\tau = \widehat{\beta} \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \tilde{v}_{it-1}}{\widehat{\sigma_w^2}}$ is not asymptotically standard normal. LLC propose to use a complicated adjustment to the standard t-test. (see (Levin *et al.* 2002) for more reference)

# Topic 5: IPS test

- IPS focus on the following generalized equation:

$$\Delta y_{it} = \delta_i y_{it-1} + \alpha_i + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \theta_t + \beta_i t + \varepsilon_{it}.$$

- IPS focus on the following hypothesis testing:

$$H_0: \ \delta_i = 0 \text{ for all i,}$$
$$H_a: \ \delta_i < 0, \ \text{ for } i = 1, 2, \ldots, n_1, \ \delta_i = 0, \ \text{ for } i = n_1 + 1, \ldots, n.$$

- Remark 1: IPS relaxes the assumption of common $\delta$.
- Remark 2: In fact, the IPS test allows for some fraction $(n_1)$ of the individual series to have unit roots.
- Remark 2: IPS assumes the errors are independently dtrsibuted across panels.

# Topic 5: IPS test in practice

- IPS test is based on a combination of the individual unit root tests for the N crosss ectional units.
- Given each $i$, the model inherently becomes a time-series model thus one can apply ADF test and obatin the $t$ statistic, $t_i$, for $i = 1, \ldots, n$.
- For $E[t_i] = \mu$ and $var(t_i) = \sigma^2$, IPS propose using the following test statistic:

$$\bar{t} = \frac{1}{n} \sum_{j=1}^{n} t_i.$$

- Under some regularity assumptions, IPS show

$$\sqrt{n}\frac{\bar{t} - \mu}{\sigma} \xrightarrow{d} N(0, 1).$$

- IPS compute $\mu$ and $\sigma^2$ by Monte Carlo simulations and report critical values for various values of $n$ and $T$.

# Topic 5: Second-generation panel unit root test

- Many panel data are cross-sectionally dependent. However, all first-generation tests do not allow for this dependent structure.
- A number of panel unit root tests are developed that allow for cross-section dependence. These tests are called "second-generation" panel unit root test.
- Among the all second-generation panel unit root tests in the literature, we will introduce the CIPS test proposed by (Pesaran 2007).

# Topic 5: CADF regression

- Pesaran considers the following dynamic panel model with cross-sectional dependence:

$$y_{it} = \gamma_i y_{it-1} + (1-\gamma_i)a_i + u_{it},$$
$$u_{it} = \phi_i f_t + \varepsilon_{it}.$$

- Therefore, we can rewrite the above model as an ADF type equation

$$\Delta y_{it} = \beta_i y_{it-1} + \alpha_i + \phi_i f_t + \varepsilon_{it}, \text{ where } \alpha_i = (1-\gamma_i)a_i.$$

- Assuming $\bar{\phi} = n^{-1}\sum_{i=1}^{n}\phi_i \neq 0$, we can write the cross-sectional mean of $y_{it}$ as

$$\bar{y}_t = \bar{y}_{t-1} + \bar{\phi} f_t + 0.$$

- Thus, Pesaran bases his test on the t-ratio of the OLS estimate $b_i$ in the "Cross Sectional Augmented Dickey Fuller" (CADF) regression,

$$\boxed{\Delta y_{it} = a_i + b_i y_{it-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it}}.$$

# Topic 5: CIPS test

- The exact null distribution of the t-ratio, $t_i(n, T)$ will depend on nuisance parameters (although not asymptotically), and Pesaran conducts simulations to derive the critical values.

- By using each t-ratio obtained from the CADF regression, Pesaran considers the following cross-sectionally augmented version of the IPS test,

$$CIPS(n, T) = \frac{1}{n} \sum_{i=1}^{T} t_i(n, T)$$

- $t_i(n, T)$ is the CADF statistic for the $i^{th}$ cross-section unit given by the t-ratio of the coefficient on $y_{it-1}$ in the CADF regression.

- Pesaran also considers an average of the truncated version of the CADF to help deal with the problems created by a lack of independence of the individual CADF statistics. (see more reference on section 4 of (Pesaran 2007)).

- To perform the LLC test,
  `xtunitroot llc varname [if] [in], [, LLC options]`
- LLC options: `trend`, `noconstant`, `demean`, `lags`, `kernel`
- To perform the IPS test,
  `xtunitroot ips varname [if] [in], [, IPS options]`
- IPS options: `trend`, `demean`, `lags`.
- To perform the CIPS test,
  `xtcips varname [, CIPS options]`
- CIPS options: `maxlags`, `bglags`, `trend`, `noc`

# Topic 5: LLC test Stata example

```
. xtunitroot llc log_GDP, lags(1) trend

Levin-Lin-Chu unit-root test for log_GDP
----------------------------------------------
Ho: Panels contain unit roots           Number of panels  =     97
Ha: Panels are stationary               Number of periods =     21

AR parameter: Common                    Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 8.00 lags average (chosen by LLC)
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 Unadjusted t        -20.7734
 Adjusted t*          -8.2236       0.0000
------------------------------------------------------------------------------
```

# Topic 5: IPS test Stata example

```
. xtunitroot ips log_GDP, lags(1) trend

Im-Pesaran-Shin unit-root test for log_GDP
-----------------------------------------------
Ho: All panels contain unit roots           Number of panels  =     97
Ha: Some panels are stationary              Number of periods =     21

AR parameter: Panel-specific                Asymptotics: T,N -> Infinity
Panel means:  Included                                       sequentially
Time trend:   Included

ADF regressions: 1 lags
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 W-t-bar             -0.3485        0.3637
------------------------------------------------------------------------------
```

# Topic 5: CIPS test Stata example

```
. xtcips log_GDP, maxl(1) bglags(2) trend

Pesaran Panel Unit Root Test with cross-sectional and first difference mean included for log_GDP
Deterministics chosen: constant & trend

Dynamics: lags criterion decision General to Particular based on F joint test

H0 (homogeneous non-stationary): bi = 0 for all i

CIPS =     -1.593          N,T = (97,21)
```

| | 10% | 5% | 1% |
|---|---|---|---|
| Critical values at | -2.51 | -2.56 | -2.66 |

# Reference

Im, Kyung So, M.Hashem Pesaran and Yongcheol Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics* **115**(1), 53–74.

Levin, Andrew, Chien-Fu Lin and Chia-Shang James Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* **108**(1), 1–24.

Pesaran, M. Hashem (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* **22**(2), 265–312.