Empirical Panel Data: Lecture 12

Instructor: Chaoyi Chen
NJE & MNB

May 23, 2023

Topic 5: Cross-sectional dependence

- Considering the standard panel-data model

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

- We wish to test if ε_{it} is correlated across cross sections or not.
- That is we are in the interest of the following hypothesis:

$$H_0 : \rho_{ij} = \rho_{ji} = cor(\varepsilon_{it},\ \varepsilon_{jt}) = 0 \text{ if } i \neq j$$
$$H_1 :\ \rho_{ij} = \rho_{ji} \neq 0 \text{ for some } i \neq j,$$

- where

$$\rho_{ij} = \rho_{ji} = \frac{\sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{jt}}{\left(\sum_{t=1}^{T} \varepsilon_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} \varepsilon_{jt}^2\right)^{1/2}}$$

Topic 5: Breusch and Pagan test

- If $n \longrightarrow \infty$ and $T \longrightarrow \infty$, we can propose using a Lagrange Multiplier Test (LM) test proposed by Breusch and Pagan (Breusch and Pagan 1980):

$$LM = T \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{\rho}_{ij}^2$$

- $\widehat{\rho}_{ij}$ is the sample estimate of the pairwise correlation of the residuals

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

- Under the null hypothesis as n and $T \longrightarrow \infty$, we have

$$LM \xrightarrow{d} \chi^2_{n(n-1)/2}.$$

Topic 5: Pesaran's CD test

- If T is fixed, BP test exhibits substantial size distortions.
- Pesaran proposed the following alternative to accomodate the large n fixed t panel (Pesaran 2004):

$$CD = \sqrt{\frac{2T}{n(n-1)}} \left(\sum_{i=1}^{n-1} \sum_{j=1\imath+1}^{n} \widehat{\rho}_{ij} \right).$$

- Under the null hypothesis as $n \longrightarrow \infty$ and T is fixed, Pesaran shows

$$CD \xrightarrow{d} N(0,1).$$

- Pesaran's CD test can also extend to unbalanced panel with a a slightly modified version of CD statitsic. See more reference on (Pesaran 2004).

Topic 5: Stata command for Pesaran's CD test

- To perform the CD test,
  1. `xtcsd, pesaran [CD options]`

     CD options:
     `abs`: computes the average absolute value of the off-diagonal elements of the crosssectional correlation matrix of residuals.
     `show`: shows the cross-sectional correlation matrix of residuals.
  2. `xtcsi, depvar indepvars [if] [in] [, trend ]`
- `xtcsd` is a postestimation command valid for use after running an FE or RE model.
- `xtcsi` also implements the error test for cross-section independence. Pros: `xtcsi` is not a postestimation comman. Cons: Only balanced panel time-series data applies to `xtcsi`.

Topic 5: Pesaran's CD test Stata example

- xtreg lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, fe
  xtcsd, pesaran

lrgdpna_gr~h	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lccon_growth	.3785448	.0067975	55.69	0.000	.3652201	.3918695
lck_growth	.0511562	.0096118	5.32	0.000	.0323149	.0699976
lpop_growth	.566651	.0534984	10.59	0.000	.461782	.6715199
_cons	.9113573	.1236405	7.37	0.000	.6689939	1.153721
sigma_u	1.1230599					
sigma_e	5.4526324					
rho	.04069581	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(179, 9046) = 1.77                Prob > F = 0.0000

. xtcsd, pesaran

Pesaran's test of cross sectional independence =    57.737, Pr = 0.0000
```

Topic 5: Panel cointegration

- Consider the following panel data model

$$y_{it} = \alpha_i + \beta_i^\top x_{it} + \varepsilon_{it}$$

- $\{y_{it}\}$ and $\{x_{it}\}$ are I(1) times series for each i.
- We are interested to test if y_{it} and x_{it} are cointegrated such that the linear combination between them forms a panel stationary ε_{it}.
- Why we care? If $\{y_{it}\}$ and $\{x_{it}\}$ are $I(1)$ and noncointegrated, ε_{it} is I(2) series too. Similar to what you learned in time series analysis, we will have a panel spurious regression.

Topic 5: Three panel cointegration tests

- Many tests have been developed since then based on different approaches and different null hypothesis. Among others, we will introduce three residual-based tests for the null of noncointegration.
  1. Kao test (Kao 1999): Assume cross-sectional independence and apply to homogeneous panel.
  2. Pedroni test (Pedroni 1999, Pedroni 2004): Assume cross-sectional independence and apply to both homogeneous and heterogeneous panel.
  3. Westerlund error correction test (Westerlund 2007): Allow cross-sectional dependence (by bootstrapping) and apply to heterogeneous panel.

Topic 5: Kao (1999) test

- Kao considers the homogeneous panel data model. The estimated homogeneous panel model from a within-group transformation is

$$y_{it} - \bar{y}_{it} = \widehat{\beta}^{\top} (x_{it} - \bar{x}_{it}) + \widehat{\varepsilon}_{it}.$$

- Therefore, under the null, $\{\widehat{\varepsilon}_{it}\}$ behave like an integrated process for large T and N. Intuitively, we can propose testing the pooled AR(1) coefficient estimator using $\widehat{\varepsilon}_{it}$. That is

$$\widehat{\varepsilon}_{it} = \rho \widehat{\varepsilon}_{it-1} + u_{it}.$$

- Therefore, the panel cointegration test becomes a generic panel unit root test.

Topic 5: Kao (1999) test statistics

- Kao proposes five test statistics based on modifying Dickey-Fuller (DF) test statistic.
  1. Unadjusted DF t: based on DF test assuming strict exogeneity and absence of serial correlation
  2. Unadjusted modified DF t: based on DF test assuming strict exogeneity and absence of serial correlation
  3. DF t: based on DF test allowing serial correlation
  4. Modified DF t: based on DF test allowing serial correlation
  5. *ADF* t: based on Augmented DF (ADF) test.
- Remark 1: Simulation results in Kao (1999) show that both DF based and ADF based test statistics have reasonable empirical size unless the errors are negatively correlated.
- Remark 2: When T is small (10 or 25), all the test statistics are subject to size distortions even when n is large. *DF* based tend to keep the nominal size better than *ADF* based.

Topic 5: Pedroni (1999, 2004) test

- Pedroni (1999, 2004) considers the heterogeneous panel regression model

$$y_{it} = \alpha_i + \beta_i^\top x_{it} + \varepsilon_{it},$$

- $\{y_{it}\}$ and $\{x_{it}\}$ are I(1) times series and satisfy conditions for the functional central limit theorem and independent over i.
- For each i, we can estimate the model by OLS and examine

$$\widehat{\varepsilon}_{it} = \rho_i \widehat{\varepsilon}_{it-1} + u_{it}.$$

- Pedroni considers a variant of IPS test statistic. For each i, Pedroni obtains the ADF or Phillips–Perron (see (Phillips and Perron 1988) for more reference) test statistic Z_i and proposes using

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i.$$

- Similar to IPS test, this statistic requires mean and variance adjustments for it to have a standard normal distribution in the limit.

Topic 5: Pedroni (1999, 2004) test statistics

- Pedroni proposes seven test statistics that use the resulting regression residuals from each i by OLS.
- Three are for heterogeneous panel (panel-specific AR parameter)
  1. Modified Phillips–Perron t
  2. Phillips–Perron t
  3. Augmented Dickey–Fuller t
- Four are for homogeneous panel (AR parameter is the same over panels)
  1. Modified variance ratio
  2. Modified Phillips–Perron t
  3. Phillips–Perron t
  4. Augmented Dickey–Fuller t
- Reamrk 1: All the proposed test statistics follow a standard normal distribution under the null.
- Remark 2: Pedroni calls the panel-specific-AR test statistics "group-mean statistics" and the same-AR test statistics "panel cointegration statistics".

Topic 5: Westerlund (2007) test

- Following (Banerjee *et al.* 1998) in the time series literature, Westerlund (2007) considers the following data-generating process

$$\Delta y_{it} = \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{it-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{it-j} + \delta_i \left(y_{it-1} + \beta_i^\top x_{it-1} \right) + \varepsilon_{it}.$$

$$\implies \Delta y_{it} = \sum_{j=1}^{p_i} \alpha_{ij} \Delta y_{it-j} + \sum_{j=-q_i}^{p_i} \gamma_{ij} \Delta x_{it-j} + \delta_i y_{it-1} + \lambda_i^\top x_{it-1} + \varepsilon_{it}.$$

- If $\delta_i = 0$, there is no error-correction term and $\{y_{it}\}$ is a unit root process that is not cointegrated with $\{x_{it}\}$.

- Given each i, we can estimate above model by OLS and obtain $\widehat{\delta}_i$. Westerlund proposes using the two group mean statistics, which defined as

$$G_\tau = \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\delta}_i}{se(\widehat{\delta}_i)}, \quad G_\alpha = \frac{1}{n} \sum_{i=1}^{n} \frac{T\widehat{\delta}_i}{\widehat{\delta}_i(1)}.$$

- Westerlund also proposes two more panel tests statistics, P_τ and P_α.

Topic 5: Westerlund (2007) test with cross sectional correlation

- Westerlund (2007) also allows for the cross sectional dependence by using a bootstrap approach (The method resembles (Chang 2004)):
  - First step: Fit the least-squares regression and obtain $\widehat{w}_t = (\varepsilon_t^\top, \Delta x_t^\top)^\top$, where ε_t and Δx_t are the observations for individual i. Then we generate bootstrap samples $w_t^* = (\varepsilon_t^{*\top}, \Delta x_t^{*\top})^\top$.
  - Second step: Using the bootstrap samples to generate Δy_{it}^* and then generate y_{it}^* and x_{it}^* under the null hypothesis.
  - Third step: Estimate the bootstrapped error-correction test of interest and obtain bootstrap test by t_1^*.
  - Fourth step: Repeat step one-three by S times and obtain $t_1^*, \ t_2^*, \ \ldots, \ t_S^*$. For a one-sided 5% nominal-level test, we then obtain the lower 5% quantile, say t_C^*, of this distribution. We reject the null hypothesis if the calculated sample value of the statistic is smaller than t_C^*.

Topic 5: Stata command for panel cointegration test

- To perform the Kao test,
  - `xtcointtest kao` depvar varlist [if] [in] [, Kao options]
  - Kao options: `lags(lspec)`, `kernel(kspec)`, `demean`
- To perform the Pedroni test,
  - `xtcointtest pedroni` depvar varlist [if] [in] [, Pedroni options]
  - Pedroni options: `ar(panelspecific|same)`, `trend`, `noconstant`, `lags(lspec)`, `kernel(kspec)`, `demean`
- To perform the Westerlund (2007) test,
  - `xtwest` depvar varlist [if] [in] [, Westerlund options]
  - Westerlund options: `lags(#[#])`, `leads(#[#])`, `lrwindow(#)`, `constant`, `trend`, `bootstrap(#)`

Topic 5: Kao test Stata example

- `xtcointtest kao lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, lags(1)`

```
Kao test for cointegration
------------------------------
H0: No cointegration                    Number of panels       =    180
Ha: All panels are cointegrated         Avg. number of periods = 49.272

Cointegrating vector: Same
Panel means:          Included          Kernel:          Bartlett
Time trend:           Not included      Lags:            2.24 (Newey-West)
AR parameter:         Same              Augmented lags:  1
--------------------------------------------------------------------------
                                        Statistic        p-value
--------------------------------------------------------------------------
 Modified Dickey-Fuller t               -95.2875         0.0000
 Dickey-Fuller t                        -69.6779         0.0000
 Augmented Dickey-Fuller t              -45.7378         0.0000
 Unadjusted modified Dickey-Fuller t    -1.7e+02         0.0000
 Unadjusted Dickey-Fuller t             -77.8970         0.0000
--------------------------------------------------------------------------
```

Topic 5: Pedroni test with panel-specific AR parameters Stata example

- `xtcointtest pedroni lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, lags(1)`

```
Pedroni test for cointegration
------------------------------
H0: No cointegration                    Number of panels       =    180
Ha: All panels are cointegrated         Avg. number of periods = 50.272

Cointegrating vector: Panel specific
Panel means:          Included          Kernel:          Bartlett
Time trend:           Not included      Lags:            3.00 (Newey-West)
AR parameter:         Panel specific    Augmented lags:  1
------------------------------------------------------------------------------
                                        Statistic        p-value
------------------------------------------------------------------------------
 Modified Phillips-Perron t             -66.0789         0.0000
 Phillips-Perron t                      -67.9837         0.0000
 Augmented Dickey-Fuller t              -67.8122         0.0000
------------------------------------------------------------------------------
```

Topic 5: Pedroni test with same AR parameters Stata example

- `xtcointtest pedroni lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, lags(1)`

```
Pedroni test for cointegration
──────────────────────────────
H0: No cointegration                        Number of panels       =    180
Ha: All panels are cointegrated             Avg. number of periods = 50.272

Cointegrating vector: Panel specific
Panel means:          Included              Kernel:           Bartlett
Time trend:           Not included          Lags:             3.00 (Newey-West)
AR parameter:         Same                  Augmented lags:   1
──────────────────────────────────────────────────────────────────────────────
                                            Statistic         p-value
──────────────────────────────────────────────────────────────────────────────
 Modified variance ratio                     23.6854          0.0000
 Modified Phillips-Perron t                 -63.9103          0.0000
 Phillips-Perron t                          -57.3284          0.0000
 Augmented Dickey-Fuller t                  -57.7981          0.0000
──────────────────────────────────────────────────────────────────────────────
```

- `xtwest lrgdpnagrowth lccongrowth lckgrowth lpopgrowth, lags(1)`

```
Results for H0: no cointegration
With 180 series and 3 covariates
```

Statistic	Value	Z-value	P-value
Gt	-4.398	-35.219	0.000
Ga	-34.000	-56.256	0.000
Pt	-64.546	-35.768	0.000
Pa	-36.354	-68.161	0.000

Reference

Banerjee, Anindya, Juan Dolado and Ricardo Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis* **19**(3), 267–283.

Breusch, T. S. and A. R. Pagan (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies* **47**(1), 239.

Chang, Yoosoon (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* **120**(2), 263–293.

Kao, Chihwa (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics* **90**(1), 1–44.

Pedroni, Peter (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* **61**(s1), 653–670.

Pedroni, Peter (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric Theory*.

Pesaran, Hashem (2004). General diagnostic tests for cross section dependence in panels. *CESifo Working Papers*.

Phillips, Peter C. and Pierre Perron (1988). Testing for a unit root in time series regression. *Biometrika* **75**(2), 335–346.

Westerlund, Joakim (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* **69**(6), 709–748.