

Econometric convergence test and club clustering using Stata

杜克锐

Email:kerrydu@sdu.edu.cn

山东大学经济研究院

August 20, 2017

Outline

- 1 Introduction
- 2 Econometric convergence test and club clustering
- 3 The PSECTA package
- 4 Monte Carlo simulation
- 5 Example



Introduction

- Convergence in economics refers to the hypothesis that all economies would eventually converge in terms of per-capita output
- Representative literature: Baumol (1986); Bernard and Durlauf (1995); Barro and Sala-i Martin (1997); Lee et al. (1997); Luginbuhl and Koopman (2004)
- Convergence analysis has also been applied in other topics, e.g., cost of living (Phillips and Sul 2007), carbon dioxide emissions (Panopoulou and Pantelidis 2009), eco-eciency (Camarero et al. 2013), house prices (Montanes and Olmos 2013), and corporate tax (Regis et al. 2015)

Introduction

- Convergence in economics refers to the hypothesis that all economies would eventually converge in terms of per-capita output
- Representative literature: Baumol (1986); Bernard and Durlauf (1995); Barro and Sala-i Martin (1997); Lee et al. (1997); Luginbuhl and Koopman (2004)
- Convergence analysis has also been applied in other topics, e.g., cost of living (Phillips and Sul 2007), carbon dioxide emissions (Panopoulou and Pantelidis 2009), eco-eciency (Camarero et al. 2013), house prices (Montanes and Olmos 2013), and corporate tax (Regis et al. 2015)

Introduction

- Phillips and Sul (2007) proposed a novel approach (termed 'log t' regression test) .
- ① Accommodates heterogeneous agent behavior and evolution in that behavior.
- ② Does not impose any particular assumptions concerning trend stationarity or stochastic nonstationarity, thereby being robust to the stationarity property of the series.

Introduction

- Another commonly concerned issue involved in the convergence analysis is the possible existence of convergence clubs.
- Traditional studies typically divided all the individuals into subgroups based on some prior information (e.g., geographical location, institution)
- Phillips and Sul (2007) constructed a new algorithm to identify clusters of convergence subgroups.

Introduction

- In this article, we introduce a new Stata module 'psecta' to perform the econometric convergence test and club clustering developed by Phillips and Sul (2007).

Time varying factor representation

$$X_{it} = g_{it} + a_{it} \quad (1)$$

where g_{it} represents systematic components such as permanent common components, and a_{it} embodies transitory components.

$$X_{it} = \left(\frac{g_{it} + a_{it}}{u_t} \right) u_t = \delta_{it} u_t \quad (2)$$

where δ_{it} is a time varying idiosyncratic element and u_t is a single common component.

Time varying factor representation

$$h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}} \quad (3)$$

h_{it} is called the relative transition parameter which measures the loading coefficient relative to the panel average at time t . Eq. (3) indicates that the cross-sectional mean of h_{it} is unity and the cross-sectional variance of h_{it} satisfies the following condition:

$$H_{it} = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 \rightarrow 0 \text{ if } \lim_{t \rightarrow \infty} \delta_{it} = \delta, \text{ for all } i.$$

The log t regression test

- The convergence of X_{it} requires the following condition:

$$\lim_{t \rightarrow \infty} \frac{X_{it}}{X_{jt}} = 1, \text{ for all } i \text{ and } j \quad (4)$$

$$\lim_{t \rightarrow \infty} \delta_{it} = \delta, \text{ for all } i \quad (5)$$

- Assume the loading coefficient δ_{it} as

$$\delta_{it} = \delta_i + \sigma_{it}\xi_{it}, \quad \sigma_{it} = \frac{\sigma_i}{L(t)t^\alpha}, \quad t \geq 1, \sigma_i > 0 \text{ for all } i$$

The log t regression test

- Phillips and Sul (2007) developed a regression t test for the null hypothesis of convergence

$$\mathcal{H}_0 : \delta_i = \delta \text{ and } \alpha \geq 0$$

- Specifically, the hypothesis test can be implemented through the following 'log t' regression model:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log(\log(t)) = a + b\log(t) + \varepsilon_t \quad (6)$$

for $t = [rT], [rT] + 1, \dots, T$ with $r > 0$

The log t regression test

- Phillips and Sul (2007) further showed that $b = 2\alpha$ and \mathcal{H}_0 is conveniently tested through the weak inequality null $\alpha \geq 0$. It implies a one-sided t test.

$$t_b = \frac{\hat{b} - b}{s_b} \Rightarrow N(0, 1)$$

Club convergence test and clustering

- ① Cross-section sorting
- ② Core group Formation
- ③ Sieve individuals for club membership
- ④ Recursion and stopping rule
- ⑤ Club merging



The logtreg command

- **logtreg** performs the log t test using linear regression with heteroskedasticity- and autocorrelation-consistent standard errors.

- Syntax

```
logtreg varname [ if ] [ in ], [ kq(#) nomata ]
```

- Options

- **kq(#)** specifies the first kq proportion of the data to be discarded before regression; default is 0.3.
- **nomata** implements the regression mainly through the Stata routines; by default, user-written mata functions are used.

The psecta command

- **psecta** implements club convergence and clustering analysis using the algorithm proposed by Phillips and Sul (2007)
- Syntax

```
psecta varname, [ name(varname) kq(#) gen(newvarname)  
cr(#) incr(#) maxcr(#) adust fr(#) nomata  
noprtlogtreg ]
```

- Options
 - **kq(#)** specifies the first kq proportion of the data to be discarded before regression; default is 0.3.
 - **cr(#)** specifies the critical value for club clustering; default is 0.
 - **incr(#)** specifies the increment of cr when the initial cr value fails to sieve individuals for clusters; default is 0.05.

The psecta command

- **psecta** implements club convergence and clustering analysis using the algorithm proposed by Phillips and Sul (2007)
- Syntax

```
psecta varname, [ name(varname) kq(#) gen(newvarname)  
cr(#) incr(#) maxcr(#) adjust fr(#) nomata  
noprtlogtreg ]
```

- Options
 - **maxcr(#)** specifies the maximum of cr value; default is 50.
 - **adjust** specifies using the adjusted method proposed by Schnurbus et al. (2016) .
 - **fr(#)** specifies sorting individuals by the time series average of the last fr proportion periods; The default is fr(0).

The psecta command

- **psecta** implements club convergence and clustering analysis using the algorithm proposed by Phillips and Sul (2007)
- Syntax

```
psecta varname, [ name(varname) kq(#) gen(newvarname)  
cr(#) incr(#) maxcr(#) adust fr(#) nomata  
noprtlogtreg ]
```

- Options
 - **maxcr(#)** specifies the maximum of cr value; default is 50.
 - **nomata** implements the regression mainly through the Stata routines; by default, user-written mata functions are used.
 - **gen(newvarname)** creates a new variable to store club classifications.

The psecta command

- **psecta** implements club convergence and clustering analysis using the algorithm proposed by Phillips and Sul (2007)

- Syntax

```
psecta varname, [ name(varname) kq(#) gen(newvarname)  
cr(#) incr(#) maxcr(#) adust fr(#) nomata  
noprtlogtreg ]
```

- Options

- **name(varname)** specifies a panel variable to be displayed for the clustering results; by default, the panel variable specified by xtset is used.
- **noprtlogtreg** suppresses the estimation results of the logtreg.

The scheckmerge command

- **scheckmerge** performs the log t test for all pairs of adjacent clubs.
- Syntax

```
scheckmerge varname, club(varname) kq(#) [ mdiv  
nomata ]
```

- Options
 - **club(varname)** specifies the initial club classifications.
 - **kq(#)** specifies the first kq proportion of the data to be discarded before regression; default is 0.3.
 - **mdiv** specifies including the divergence group for the log t test; by default, the divergence group is excluded.
 - **nomata** implements the regression mainly through the Stata routines; by default, user-written mata functions are used.

The `imergeclub` command

- **imergeclub** iteratively conducts merging adjacent clubs.
- Syntax

```
imergeclub varname, club(varname) kq(#) [ name(varname)  
gen(newvarname) imore mdiv nomata noprtlogtreg ]
```

- Options
 - **club(varname)** specifies the initial club classifications.
 - **kq(#)** specifies the first kq proportion of the data to be discarded before regression; default is 0.3.
 - **mdiv** specifies including the divergence group for the log t test; by default, the divergence group is excluded.
 - **nomata** implements the regression mainly through the Stata routines; by default, user-written mata functions are used.

The `imergeclub` command

- **imergeclub** iteratively conducts merging adjacent clubs.
- Syntax

```
imergeclub varname, club(varname) kq(#) [ name(varname)  
gen(newvarname) imore mdiv nomata noprtlogtreg ]
```

- Options
 - **imore** specifies merging clubs iteratively until no clubs can be merged.
 - **gen(newvarname)** creates a new variable to store the new club classifications.
 - **name(varname)** specifies a panel variable to be displayed for the clustering results.
 - **noprtlogtreg** suppresses the estimation results of the logtreg.

The pfilter command

- **pfilter** applies the **[TS] tfilter** command into panel data context. It can be used to extract trend and cyclical components for each individual in the panel, respectively
- Syntax

```
pfilter varname, method(string) [ trend(newvarname)  
cyc(newvarname) options ]
```

- Options
 - **method(string)** specifies the filter method; it should be chosen from {bk, bw, cf, hp}.
 - **trend(newvarname)** creates a new variable for the trend component.
 - **cyc(newvarname)** creates a new variable for the cyclical component..
 - **options** are any options available for tfilter. (See [TS] tfilter).

Data generating process

$$X_{it} = \theta_{it}\mu_t, \theta_{it} = \theta_i + \theta_{it}^0,$$
$$\theta_{it}^0 = \rho_i\theta_{it-1}^0 + \varepsilon_{it}$$

where $t = 1, \dots, T$; $\varepsilon_{it} \sim iid N(0, \sigma_i^2 \log(t+1)^{-2} t^{-2\alpha_i})$, $\sigma_i \sim U[0.02, 0.28]$, $\alpha_i \sim U[0.2, 0.9]$; $\rho_i \sim U[0, 0.4]$. Note that the common component μ_t cancels out in the test procedure.

Data generating process

- *Case 1*: One convergence club and one divergence subgroup. We consider two equal sized groups in the panel with numbers $N_1 = N_2 = \frac{N}{2}$. We set $\theta_i = 1$ and $\theta_i \sim U[1.5, 5]$ for the first and second groups, respectively. It implies that the first group forms a convergence club and the second group is divergent.
- *Case 2*: Two convergence clubs. Two groups are set as in *Case 1* except that $\theta_i = 1$ and $\theta_i = 2$ for the first and second groups, respectively.

Data generating process

- *Case 3*: We collect per capita GDP of the United States and Democratic Republic of the Congo, and denote them as X_t^U and X_t^C , respectively¹. The simulation data is generated by $\frac{N}{2}$ copies of X_t^U and X_t^C with noises as follows.

$$X_{it}^j = X_t^j + \theta_{it}^0 X_t^j, j = \{U, C\}$$

where θ_{it}^0 is set as described above except that $\alpha_j = (0.1, 0.3, 0.6, 0.8)$.

¹The period is 1950-2014, namely, $T = 65$.

Simulation Results

Table: Simulation result of Case 1

T \ N	40	80	120
20	0.097	0.070	0.082
40	0.048	0.035	0.047
60	0.023	0.032	0.038
100	0.027	0.033	0.039

Simulation Results

Table: Simulation result of Case 2

T \ N	40	80	120
20	0.026	0.023	0.046
40	0.014	0.031	0.056
60	0.012	0.022	0.040
100	0.015	0.015	0.042

Simulation Results

Table: Simulation result of Case 3

$\alpha_i \backslash N$	40	80	120
0.1	0.034	0.054	0.044
0.3	0.013	0.030	0.044
0.6	0.014	0.030	0.041
0.8	0.010	0.031	0.040

Example

- The example provided here is a replication of the key results of Phillips and Sul (2009). They collected a panel data covering 152 economies during the period of 1970-2003 from PWT. They first examined whether the convergence hypothesis holds for the whole sample. Then they investigated the possibility of club convergence using their proposed clustering algorithm. The replication is conducted as follows.

- <https://sites.google.com/site/kerrydu2016/home/stata-files>
- <http://kerrydu.weebly.com/computer-code.html>



