Econometric convergence test and club clustering using Stata

Kerui Du
The Center for Economic Research, Shandong University
Jinan, Shandong/China
kerrydu@sdu.edu.cn

Abstract. In this article, I introduce a new Stata module including five commands to perform econometric convergence analysis and club clustering proposed by Phillips and Sul (2007, Econometrica 75(6): 1771-1855). The **logtreg** command performs the log t regression test. The **psecta** command implements the clustering algorithm to identify convergence clubs. The **scheckmerge** command conducts the log t regression test for all pairs of adjacent clubs. The **imergeclub** command tries to iteratively merge adjacent clubs. The **pfilter** command extracts trend/cyclical component of a time series of each individual respectively in panel data. An example from Phillips and Sul (2009, Journal of Applied Econometrics 24(7): 1153-1185) is also provided to illustrate the use of these commands. Additionally, I use Monte Carlo simulations to exemplify the effectiveness of the clustering algorithm.

Keywords: logtreg, psecta, scheckmerge, imergeclub, pfilter, convergence, club clustering, log t test

1 Introduction

Convergence in economics refers to the hypothesis that all economies would eventually converge in terms of per-capita output. This issue has played a central role in the empirical growth literature (Pesaran 2007). A large body of literature (e.g., Baumol 1986; Bernard and Durlauf 1995; Barro and Sala-i Martin 1997; Lee et al. 1997; Luginbuhl and Koopman 2004) has contributed to developing methods for convergence tests and empirically investigating the convergence hypothesis across different countries and regions. In the past years, convergence analysis has also been applied in other topics such as cost of living (Phillips and Sul 2007), carbon dioxide emissions (Panopoulou and Pantelidis 2009), eco-efficiency (Camarero et al. 2013), house prices (Montañés and Olmos 2013), and corporate tax (Regis et al. 2015), etc.

Phillips and Sul (2007) proposed a novel approach (termed 'log t' regression test [1]) to test the convergence hypothesis based on a nonlinear time-varying factor model. The proposed approach has the following merits. First, it accommodates heterogeneous agent behavior and evolution in that behavior. Second, the proposed test does not impose any particular assumptions concerning trend stationarity or stochastic nonstationarity, thereby being robust to the stationarity property of the series. Phillips and Sul (2009) showed that the traditional

[1] It is also called "log t test" for short.

convergence tests for economic growth have some pitfalls. For instance, estimation of augmented Solow regression under transitional heterogeneity is biased and inconsistent due to the issues of omitted variables and endogeneity. Conventional cointegration tests typically have low power to detect the asymptotic co-movement since the existence of a unit root in the differential of the series does not necessarily lead to the divergence conclusion.

Another commonly concerned issue involved in the convergence analysis is the possible existence of convergence clubs. Regarding this issue, traditional studies typically divided all the individuals into subgroups based on some prior information (e.g., geographical location, institution), and then tested the convergence hypothesis for each subgroup respectively. Phillips and Sul (2007) constructed a new algorithm to identify clusters of convergence subgroups. The developed algorithm is a data-driven method which avoids ex-ante sample separation. The relative transition parameter mechanism which Phillips and Sul (2007) proposed to characterize individual variations fits in with some commonly used models [2]. It can be used as a general panel method to cluster individuals into groups with similar transition paths.

In practice, Phillips and Sul (2007; 2009) provided GAUSS codes to perform their empirical studies. Recently, Schnurbus et al. (2016) provided a set of R functions to replicate the key results of Phillips and Sul (2009). 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).

The remainder of the paper is organized as follows: Section 2 briefly describes the methodology of Phillips and Sul (2007); Sections 3-7 explain the syntax and options of the new commands; Section 8 performs Monte Carlo simulations to examine the effectiveness of the clustering algorithm; Section 9 illustrates the use of the commands with an example from Phillips and Sul (2009).

2 Econometric convergence test and club clustering

2.1 Time varying factor representation

The start point of the model is decomposing the panel data X_{it} as:

$$X_{it} = g_{it} + a_{it} \tag{1}$$

where g_{it} represents systematic components such as permanent common components, and a_{it} embodies transitory components. To separate common components from idiosyncratic components, Eq. (1) is further transformed as:

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

where δ_{it} is a time varying idiosyncratic element and u_t is a single common component. Eq. (2) is a dynamic factor model where u_t captures some deter-

[2] See, for example, economic growth with heterogeneous technological progress (Parente and Prescott 1994; Howitt and Mayer-Foulkes 2005), income processes with heterogeneity (Baker 1997; Moffitt and Gottschalk 2002), and stock price factor model (Menzly et al. 2002; Ludvigson and Ng 2007), etc.

ministic or stochastically trending behavior, and the time varying factor-loading coefficient δ_{it} measures the idiosyncratic distance between the common trend component u_t and X_{it}.

In general, we cannot directly estimate the model without imposing some restrictions on δ_{it} and u_t. Thus, Phillips and Sul (2007) proposed removing the common factor as follows.

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

h_{it} is called the relative transition parameter which measures the loading coefficient relative to the panel average at time t. In other words, h_{it} traces out a transition path of individual i in relation to the panel average. 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:

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

2.2 The log t regression test

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

$$\lim_{t\to\infty} \frac{X_{it}}{X_{jt}} = 1, \ for \ all \ i \ and \ j \tag{4}$$

Phillips and Sul (2007) defined this condition as the relative convergence. It is equivalent to the convergence of the time varying factor-loading coefficient

$$\lim_{t\to\infty} \delta_{it} = \delta, \ for \ all \ i \tag{5}$$

Assume the loading coefficient δ_{it} as

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

where $L(t)$ is a slowly varying function. Possible choices for $L(t)$ can be $log(t)$, $log^2(t)$, or $log(log(t))$. The Monte Carlo simulations in Phillips and Sul (2007) indicate that $L(t) = log(t)$ produces the least size distortion and the best test power. Thus, we set $L(t) = log(t)$ in our Stata codes.

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

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

against the alternative $\mathcal{H}_A : \delta_i \neq \delta \ or \ \alpha < 0$. Specifically, the hypothesis test can be implemented through the following 'log t' regression model:

$$\begin{aligned} &log\left(\frac{H_1}{H_t}\right) - 2log(log(t)) = a + blog(t) + \varepsilon_t \\ &for \ t = [rT], [rT]+1, ..., T \ with \ r > 0 \end{aligned} \tag{6}$$

The selection of the initiating sample fraction r might influence the results of the above regression. The Monte Carlo experiments indicate that $r \in [0.2, 0.3]$ achieves a satisfactory performance. Specifically, it is suggested to set $r = 0.3$ for the small or moderate $T(\leq 50)$ sample and set $r = 0.2$ for the large $T(\geq 100)$ sample.

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. Under some technical assumptions, the limit distribution of the regression t statistic is

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

where $s_b^2 = l\hat{var}(\hat{\varepsilon}_t)\left[\sum_{t=[rT]}^{T}\left(log(t) - \frac{1}{T-[rT]+1}\sum_{t=[rT]}^{T} log(t)\right)^2\right]^{-1}$ and $l\hat{var}(\hat{\varepsilon}_t)$ is a conventional HAC estimate formed from the regression residuals.

2.3 Club convergence test and clustering

Rejection of the null hypothesis of convergence for the whole panel cannot rule out the existence of convergence in subgroups of the panel. To investigate the possibility of convergence clusters, Phillips and Sul (2007) developed a data-driven algorithm. Schnurbus et al. (2016) advocated making some minor adjustments to the original algorithm. We provide a sketch of their ideas as follows.

(1) **Cross-section sorting**.

Sort individuals in the panel decreasingly according to their observations in the last period. If there is substantial time series volatility in the data, the sorting can be implemented based on the time series average of the last fraction (e.g., 1/2,1/3) of the sample. Index individuals with their orders $\{1, ..., N\}$.

(2) **Core group Formation**.

(2.1) Find the first k such that the test statistic of the log t regression $t_k > -1.65$ for the subgroup with individuals $\{k, k+1\}$. If there is no k satisfying $t_k > -1.65$, exit the algorithm, and conclude that there are no convergence subgroups in the panel.

(2.2) Start with the k identified in Step (2.1), perform log t regression for the subgroups with individuals $\{k, k+1, ..., k+j\}, j \in \{1, ..., N-k\}$. Choose j^* such that the subgroup with individuals $\{k, k+1, ..., k+j^*\}$ yields the highest value of the test statistic. Individuals $\{k, k+1, ..., k+j^*\}$ form a core group.

(3) **Sieve individuals for club membership**.

(3.1) Form a complementary group G_{j*}^c with all the remaining individuals not included in the core group. Add one individual from G_{j*}^c at each time to the core group and run the log t test. Include the individual in the club candidate group if the test statistic is greater than the critical value c^* [3].

(3.2) Run the log t test for the club candidate group identified in Step (3.1). If the test statistic $\hat{t}_b$ is greater than -1.65, the initial convergence club is

[3]When T is small, the sieve criterion c^* can be set to zero to ensure that it is highly conservative, whereas for large T, c^* can be set to the asymptotic 5% critical value -1.65.

obtained. If not, Phillips and Sul (2007) advocated raising the critical value c^* and repeating Steps (3.1) and (3.2) until $\hat{t}_b > -1.65$. Schnurbus et al. (2016) proposed adjusting this step as follows. If the convergence hypothesis does not hold for the club candidate group, sort the club candidates w.r.t. decreasing $\hat{t}_b$ obtained in Step (3.1). If there is some $\hat{t}_b > -1.65$, add the individual with the highest value of $\hat{t}_b$ to form an extended core group. Add one individual from the remaining candidates at a time and run the log t test and denote the test statistic $\hat{t}_b$. If the highest value of $\hat{t}_b$ is not greater than -1.65, stop the procedure, and the extended core group forms an initial convergence club. Otherwise, repeat the above procedure to add the individual with the highest $\hat{t}_b$.

(4) **Recursion and stopping rule**.

Form a subgroup of the remaining individuals which are not sieved by Step (3). Perform the log t test for this subgroup. If the test statistic is greater than -1.65, the subgroup forms another convergence club. Otherwise, repeat Steps (1)-(3) on this subgroup.

(5) **Club merging**.

Perform the log t test for all pairs of the subsequent initial clubs. Merge those clubs fulfilling the convergence hypothesis jointly. Schnurbus et al. (2016) advocated conducting club merging iteratively as follows: run the log t test for the initial Clubs 1 and 2; if they fulfill the convergence hypothesis jointly, merge them to form the new Club 1, and then come to run the log t test for the new Club 1 and the initial Club 3 jointly; if not, come to run the log t test for initial Clubs 2 and 3, so on and so forth. The new club classifications would be obtained by the above procedure. After that, one can also repeat the procedure on the newly obtained club classifications until no clubs can be merged anymore, which leads to the classifications with the smallest number of convergence clubs.

3 The logtreg command

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

3.1 Syntax

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

3.2 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.

3.3 Stored results

`logtreg` stores the following in `e()`:

Scalars	
`e(N)`	number of individuals
`e(T)`	number of time periods
`e(nreg)`	number of observations used for the regression
`e(beta)`	log t coefficient
`e(tstat)`	t statistic for log t
Matrix	
`e(res)`	table of estimation results
Macros	
`e(cmd)`	`logtreg`
`e(cmdline)`	command as typed
`e(varlist)`	name of the variable

4 The psecta command

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

4.1 Syntax

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

4.2 Options

Regression

`kq(#)` specifies the first *kq* proportion of the data to be discarded before regression; default is 0.3.

Algorithm

`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.

`maxcr(#)` specifies the maximum of cr value; default is 50.

`adjust` specifies using the adjusted method proposed by Schnurbus et al. (2016) instead of raising cr when the initial cr value fails to sieve individuals for clusters. See Schnurbus et al. (2016) for more details.

`fr(#)` specifies sorting individuals by the time series average of the last *fr* proportion periods; The default is fr(0), sorting individuals according to the last period.

`nomata` uses Stata routines; by default, user-written mata functions are used.

Variable generation

`gen(`*newvarname*`)` creates a new variable to store club classifications. For the individuals that are not classified into any convergence club, missing values are generated.

Miscellaneous

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.

4.3 Stored results

psecta stores the following in e():

Scalar

e(nclub)	number of convergent clubs

Matrix

e(bm)	log t coefficients
e(tm)	t statistics
e(club)	club classifications

Macros

e(cmd)	psecta
e(cmdline)	command as typed
e(varlist)	name of the variable

4.4 Dependency of psecta

psecta depends on the Mata function mm_which(). If not already installed, you can install it by typing ssc install moremata.

5 The scheckmerge command

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

5.1 Syntax

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

5.2 Options

Specification

club(*varname*) specifies the initial club classifications; it is required.

Regression

kq(#) specifies the first *kq* proportion of the data to be discarded before regression; it is required.

mdiv specifies including the divergence group for the log t test; by default, the divergence group is excluded.

nomata uses Stata routines; by default, user-written mata functions are used.

5.3 Stored results

scheckmerge stores the following in e():

Matrix	
`e(bm)`	log t coefficients
`e(tm)`	t statistics
Macros	
`e(cmd)`	`scheckmerge`
`e(cmdline)`	command as typed
`e(varlist)`	name of the variable

6 The imergeclub command

`imergeclub` iteratively conducts merging adjacent clubs.

6.1 Syntax

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

6.2 Options

Specification

`club(`*varname*`)` specifies the initial club classifications; it is required.

Regression

`kq(`#`)` specifies the first *kq* proportion of the data to be discarded before regression; it is required.

`mdiv` specifies including the divergence group during club merging; by default, the divergence group is excluded.

`nomata` uses Stata routines; by default, user-written mata functions are used.

Algorithm

`imore` specifies merging clubs iteratively until no clubs can be merged. By default, the procedure is conducted as follows. First, run the log t test for the individuals belonging to the initial Clubs 1 and 2 (obtained from club clustering). Second, if they fulfill the convergence hypothesis jointly, merge them to be the new Club 1, and then come to run the log t test for the new Club 1 and the initial Club 3; if not, come to run the log t test for initial Clubs 2 and 3, so on and so forth. If imore is chosen, the above procedure is repeated until no clubs can be merged.

Variable generation

`gen(`*newvarname*`)` creates a new variable to store the new club classifications. For the individuals which are not classified into any convergence club, missing values are generated.

Miscellaneous

`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.

6.3 Stored results

`scheckmerge` stores the following in `e()`:

Matrix	
`e(bm)`	log t coefficients
`e(tm)`	t statistics
Macros	
`e(cmd)`	`imergeclub`
`e(cmdline)`	command as typed
`e(varlist)`	name of the variable

7 The pfilter command

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

7.1 Syntax

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

7.2 Options

Method

`method(`*string*`)` specifies the filter method; it should be chosen from {bk, bw, cf, hp}; it is required.

Variable generation

`trend(`*newvarname*`)` creates a new variable for the trend component.

`cyc(`*newvarname*`)` creates a new variable for the cyclical component.

Miscellaneous

options are any options available for tsfilter (See [TS] **tsfilter**).

7.3 Stored results

`pfilter` stores the following in `r()`:

Macros	
`r(cmd)`	`pfilter`
`r(cmdline)`	command as typed
`r(varlist)`	name of the variable

8 Monte Carlo simulation

Phillips and Sul (2007) performed extensive simulations for the power and size of the 'log t' test. They also showed how the clustering procedure works through Monte Carlo experiments. In this section, we further do simulation exercises to exemplify the effectiveness of the clustering algorithm in the finite sample.

Referring to Phillips and Sul (2007), the following data generating process is considered.

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

where $t = 1, ..., T$; $\varepsilon_{it} \sim iid\ N(0, \sigma_i^2 log(t+1)^{-2}t^{-2\alpha_i}), \sigma_i \sim U[0.02, 0.28], \alpha_i \sim U[0.2, 0.9]$; $\rho_i \sim U[0, 0.4]$. Note that the common component μ_t cancels out in the test procedure. It is not needed to generate the realizations of μ_t.

For simulations, we set $T = 20, 40, 60, 100$ and $N = 40, 80, 120$. The number of replications is 1000. We first consider the following two cases.

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.

Tables 1 and 2 report the false discovery rate (the ratio of the replications that fail to identify the club memberships) for *Case 1* and *Case 2*, respectively. Generally speaking, the false discovery rate is acceptable. In *Case 1*, the false discovery rate is lower than 10% when $T = 20$ and it becomes lower than 5% when $T \geq 40$. In *Case 2*, the false discovery rate is lower than 5% for all combinations of N and T except for $(N = 120, T = 40)$.

Table 1: 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

Table 2: 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

For the experiments of *Case 1* and *Case 2*, the values of θ_i are given. Here we provide another experiment in which the values of θ_i are unknown, and the data are generated by copying actual data with noises. The experiment is described as follows.

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, respectively [4]. 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 $\alpha_i = (0.1, 0.3, 0.6, 0.8)$.

The result is given in Table 3. It is shown that the false discovery rate is lower than 5% for all combinations of N and α_i except for $(N = 80, \alpha_i = 0.1)$. Taking the results presented in Tables 1, 2 and 3 together, we can conclude that the clustering algorithm generally achieves a satisfactory performance.

Table 3: Simulation result of Case 3

α_i \ 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

9 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.

```
. use ps2009
(PWT152, from Phillps and Sul (2009) in Jounarl of Applied Econometrics)
. egen id=group(country)
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1970 to 2003
                delta:  1 unit
. gen lnpgdp=ln(pgdp)
. pfilter lnpgdp, method(hp) trend(lnpgdp2) smooth(400)
```

First, the pfilter command is used to wipe out the cyclical component. A new variable lnpgdp2 is generated to store the trend component. We then run the log t regression for the convergence test. The output reports the coefficient, standard error and t statistic for log(t). Since the value of the t statistic (calculated as -159.55) is less than -1.65, the null hypothesis of convergence is rejected at the 5% level.

[4]The period is 1950-2014, namely, $T = 65$.

```
. logtreg lnpgdp2,  kq(0.333)
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.8748	0.0055	-159.5544

```
The number of individuals is 152.
The number of time periods is 34.
The first 11 periods are discarded before regression.
```

Furthermore, identifying convergence clubs is conducted by psecta command. The output presents the club classifications. The noprtlogtreg option suppresses the estimation details. If not, the results of the log t regression would be displayed following each club. We put all the estimation results together in the matrix *result1* and display it. Additionally, a new variable *club* is generated to store the club classifications.

```
. psecta lnpgdp2, name(country) kq(0.333) gen(club) noprt

xxxxxxxxxxxxxxxxxxxxxxxxxxx Club classifications xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
------------------------------- Club 1 :(50)-------------------------------
 | Antigua | Australia | Austria | Belgium | Bermuda | Botswana |
 | Brunei | Canada | Cape Verde | Chile | China | Cyprus | Denmark |
 | Dominica | Equatorial Guinea | Finland | France | Germany |
 | Hong Kong | Iceland | Ireland | Israel | Italy | Japan |
 | Korea, Republic of | Kuwait | Luxembourg | Macao | Malaysia |
 | Maldives | Malta | Mauritius | Netherlands | New Zealand |
 | Norway | Oman | Portugal | Puerto Rico | Qatar | Singapore |
 | Spain | St. Kitts & Nevis | St.Vincent & Grenadines | Sweden |
 | Switzerland | Taiwan | Thailand | United Arab Emirates |
 | United Kingdom | United States |
--------------------------------------------------------------------------
------------------------------- Club 2 :(30)-------------------------------
 | Argentina | Bahamas | Bahrain | Barbados | Belize | Brazil |
 | Colombia | Costa Rica | Dominican Republic | Egypt | Gabon |
 | Greece | Grenada | Hungary | India | Indonesia | Mexico |
 | Netherlands Antilles | Panama | Poland | Saudi Arabia |
 | South Africa | Sri Lanka | St. Lucia | Swaziland | Tonga |
 | Trinidad &Tobago | Tunisia | Turkey | Uruguay |
--------------------------------------------------------------------------
------------------------------- Club 3 :(21)-------------------------------
 | Algeria | Bhutan | Cuba | Ecuador | El Salvador | Fiji |
 | Guatemala | Iran | Jamaica | Lesotho | Micronesia, Fed. Sts. |
 | Morocco | Namibia | Pakistan | Papua New Guinea | Paraguay |
 | Peru | Philippines | Romania | Suriname | Venezuela |
--------------------------------------------------------------------------
------------------------------- Club 4 :(24)-------------------------------
 | Benin | Bolivia | Burkina Faso | Cameroon | Cote d Ivoire |
 | Ethiopia | Ghana | Guinea | Honduras | Jordan | Korea, Dem. Rep. |
 | Laos | Mali | Mauritania | Mozambique | Nepal | Nicaragua | Samoa |
 | Solomon Islands | Syria | Tanzania | Uganda | Vanuatu | Zimbabwe |
--------------------------------------------------------------------------
------------------------------- Club 5 :(14)-------------------------------
 | Cambodia | Chad | Comoros | Congo, Republic of | Gambia, The |
 | Iraq | Kenya | Kiribati | Malawi | Mongolia | Nigeria |
 | Sao Tome and Principe | Senegal | Sudan |
--------------------------------------------------------------------------
```

```
------------------------------- Club 6 :(11)-------------------------------
 | Afghanistan | Burundi | Central African Republic |
 | Guinea-Bissau | Madagascar | Niger | Rwanda | Sierra Leone |
 | Somalia | Togo | Zambia |
--------------------------------------------------------------------------
------------------------------- Club 7 :(2)-------------------------------
 | Congo, Dem. Rep. | Liberia |
--------------------------------------------------------------------------
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
. mat b=e(bm)
. mat t=e(tm)
. mat result1=(b \ t)
. matlist result1, border(rows) rowtitle("log(t)") format(%9.3f) left(4)
```

log(t)	Club1	Club2	Club3	Club4	Club5	Club6	Club7
Coeff	0.382	0.240	0.110	0.131	0.190	1.003	-0.470
T-stat	9.282	6.904	3.402	2.055	1.701	6.024	-0.559

Finally, we use the scheckmerge and imergeclub commands to perform possible club merging. It is shown that the initial Clubs 4 and 5 can be merged to form a larger convergent club. The results obtained here are the same as those in Phillips and Sul (2009) and Schnurbus et al. (2016).

```
. scheckmerge lnpgdp2,  kq(0.333) club(club) mdiv

          The log t test for   Club 1+2
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.0507	0.0232	-2.1909

```
The number of individuals is 80.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------

          The log t test for   Club 2+3
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.1041	0.0159	-6.5339

```
The number of individuals is 51.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------

          The log t test for   Club 3+4
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.1920	0.0379	-5.0684

```
The number of individuals is 45.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------
```

```
         The log t test for   Club 4+5
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.0443	0.0696	-0.6360

```
The number of individuals is 38.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------
```

```
         The log t test for   Club 5+6
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-0.2397	0.0612	-3.9178

```
The number of individuals is 25.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------
```

```
         The log t test for   Club 6+7
log t test:
```

Variable	Coeff	SE	T-stat
log(t)	-1.1163	0.0602	-18.5440

```
The number of individuals is 13.
The number of time periods is 34.
The first 11 periods are discarded before regression.
------------------------------------------------------
```

```
. mat b=e(bm)
. mat t=e(tm)
. mat result2=(b \ t)
. matlist result2, border(rows) rowtitle("log(t)") format(%9.3f) left(4)
```

log(t)	Club1+2	Club2+3	Club3+4	Club4+5	Club5+6	Club6+7
Coeff	-0.051	-0.104	-0.192	-0.044	-0.240	-1.116
T-stat	-2.191	-6.534	-5.068	-0.636	-3.918	-18.544

```
. imergeclub lnpgdp2, name(country) kq(0.333) club(club) gen(finalclub) noprt
xxxxxxxxxxxxxxxxxxxxxxxxxx Final Club classifications xxxxxxxxxxxxxxxxxxxxxxxxxxxxx
------------------------------- Club 1 :(50)-------------------------------
 | Antigua | Australia | Austria | Belgium | Bermuda | Botswana |
 | Brunei | Canada | Cape Verde | Chile | China | Cyprus | Denmark |
 | Dominica | Equatorial Guinea | Finland | France | Germany |
 | Hong Kong | Iceland | Ireland | Israel | Italy | Japan |
 | Korea, Republic of | Kuwait | Luxembourg | Macao | Malaysia |
 | Maldives | Malta | Mauritius | Netherlands | New Zealand |
```

```
 | Norway | Oman | Portugal | Puerto Rico | Qatar | Singapore |
 | Spain | St. Kitts & Nevis | St.Vincent & Grenadines | Sweden |
 | Switzerland | Taiwan | Thailand | United Arab Emirates |
 | United Kingdom | United States |
--------------------------------------------------------------------------
------------------------------- Club 2 :(30)-------------------------------
 | Argentina | Bahamas | Bahrain | Barbados | Belize | Brazil |
 | Colombia | Costa Rica | Dominican Republic | Egypt | Gabon |
 | Greece | Grenada | Hungary | India | Indonesia | Mexico |
 | Netherlands Antilles | Panama | Poland | Saudi Arabia |
 | South Africa | Sri Lanka | St. Lucia | Swaziland | Tonga |
 | Trinidad &Tobago | Tunisia | Turkey | Uruguay |
--------------------------------------------------------------------------
------------------------------- Club 3 :(21)-------------------------------
 | Algeria | Bhutan | Cuba | Ecuador | El Salvador | Fiji |
 | Guatemala | Iran | Jamaica | Lesotho | Micronesia, Fed. Sts. |
 | Morocco | Namibia | Pakistan | Papua New Guinea | Paraguay |
 | Peru | Philippines | Romania | Suriname | Venezuela |
--------------------------------------------------------------------------
------------------------------- Club 4 :(38)-------------------------------
 | Benin | Bolivia | Burkina Faso | Cambodia | Cameroon | Chad |
 | Comoros | Congo, Republic of | Cote d Ivoire | Ethiopia |
 | Gambia, The | Ghana | Guinea | Honduras | Iraq | Jordan | Kenya |
 | Kiribati | Korea, Dem. Rep. | Laos | Malawi | Mali |
 | Mauritania | Mongolia | Mozambique | Nepal | Nicaragua |
 | Nigeria | Samoa | Sao Tome and Principe | Senegal |
 | Solomon Islands | Sudan | Syria | Tanzania | Uganda | Vanuatu |
 | Zimbabwe |
--------------------------------------------------------------------------
------------------------------- Club 5 :(11)-------------------------------
 | Afghanistan | Burundi | Central African Republic |
 | Guinea-Bissau | Madagascar | Niger | Rwanda | Sierra Leone |
 | Somalia | Togo | Zambia |
--------------------------------------------------------------------------
------------------------------- Club 6 :(2)-------------------------------
 | Congo, Dem. Rep. | Liberia |
--------------------------------------------------------------------------
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
. mat b=e(bm)
. mat t=e(tm)
. mat result3=(b \ t)
. matlist result3, border(rows) rowtitle("log(t)") format(%9.3f) left(4)
```

log(t)	Club1	Club2	Club3	Club4	Club5	Club6
Coeff	0.382	0.240	0.110	-0.044	1.003	-0.470
T-stat	9.282	6.904	3.402	-0.636	6.024	-0.559

10 Acknowledgments

This paper is supported by National Nature Science Foundation of China (No.71603148), China Postdoctoral Science Foundation (No.2016M590627; No.2017T100480) and Shandong Provincial Natural Science Foundation, China (No.ZR2016GB10). The Stata module introduced here benefits a lot from Donggyu Sul's GAUSS codes. I would like to thank Donggyu Sul for making his codes available to us. I greatly appreciate H. Joseph Newton (the editor) and an anonymous referee for their helpful comments and suggestions which leads to an improved version of this paper and the Stata module. I thank Min Deng for her linguistic assistance during the revision of this manuscript.

11 References

Baker, M. 1997. Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics* 15(2): 338–375.

Barro, R. J., and X. Sala-i Martin. 1997. Technological diffusion, convergence, and growth. *Journal of Economic Growth* 2: 1–26.

Baumol, W. J. 1986. Productivity growth, convergence, and welfare: what the long-run data show. *The American Economic Review* 76(5): 1072–1085.

Bernard, A. B., and S. N. Durlauf. 1995. Convergence in international output. *Journal of Applied Econometrics* 10(2): 97–108.

Camarero, M., J. Castillo, A. J. Picazo-Tadeo, and C. Tamarit. 2013. Eco-efficiency and convergence in OECD countries. *Environmental and Resource Economics* 55(1): 87–106.

Howitt, P., and D. Mayer-Foulkes. 2005. R&D, Implementation, and Stagnation: A Schumpeterian Theory of Convergence Clubs. *Journal of Money, Credit and Banking* 147–177.

Lee, K., M. H. Pesaran, R. P. Smith, et al. 1997. Growth and convergence in a multi-country empirical stochastic Solow model. *Journal of Applied Econometrics* 12(4): 357–392.

Ludvigson, S. C., and S. Ng. 2007. The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics* 83(1): 171–222.

Luginbuhl, R., and S. J. Koopman. 2004. Convergence in European GDP series: a multivariate common converging trend–cycle decomposition. *Journal of Applied Econometrics* 19(5): 611–636.

Menzly, L., T. Santos, and P. Veronesi. 2002. The time series of the cross section of asset prices. *NBER Working Paper No.9217* .

Moffitt, R. A., and P. Gottschalk. 2002. Trends in the transitory variance of earnings in the United States. *The Economic Journal* 112(478): C68–C73.

Montañés, A., and L. Olmos. 2013. Convergence in US house prices. *Economics Letters* 121(2): 152–155.

Panopoulou, E., and T. Pantelidis. 2009. Club convergence in carbon dioxide emissions. *Environmental and Resource Economics* 44(1): 47–70.

Parente, S. L., and E. C. Prescott. 1994. Barriers to technology adoption and development. *Journal of Political Economy* 102(2): 298–321.

Pesaran, M. H. 2007. A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* 138(1): 312–355.

Phillips, P. C., and D. Sul. 2007. Transition modeling and econometric convergence tests. *Econometrica* 75(6): 1771–1855.

———. 2009. Economic transition and growth. *Journal of Applied Econometrics* 24(7): 1153–1185.

Regis, P. J., J. C. Cuestas, and Y. Chen. 2015. Corporate tax in Europe: Towards convergence? *Economics Letters* 134: 9–12.

Schnurbus, J., H. Haupt, and V. Meier. 2016. Economic Transition and Growth: A Replication. *Journal of Applied Econometrics* forthcoming.

About the authors

Kerui Du is an assistant professor at the Center for Economic Research, Shandong University. His primary research interests are applied econometrics, energy and environmental economics.