Supplementary Material

Methods

Phillips and Sul (2007) [19] and Phillips and Sul (2009) [20] propose a modification in the data decomposition of the variable under study. If the data is usually broken down in the following way where the variable under study is X:

$$X_{it} = g_{it} + a_{it}$$

this is defined by g_{it} as the systemic part that affects the entire variable and includes the common component, and a_{it} as the transitory component in time. If you want to have the heterogeneity of time in the variable under study, the resulting equation is:

$$X_{it} = \left(\frac{g_{it}+a_{it}}{\mu_t}\mu_t\right) = b_{it}\mu_t,$$

Where b_{it} is the element that considers time and changes with it, includes a random component that absorbs a_{it}, and μt is the factor common to them. With this proposed factorial formulation that separates the proper parts of the object from the common ones, b_{it} is the path to a common state, determined by μ_t.

b_{it} estimation is essential for checking that different objects converge; however, this estimation is not possible without first setting some additional restrictions. In this way, Phillips and Sul (2007) [19] and Phillips and Sul (2009) [20] establish the following formulation that was the suitable relative transition path to calculate directly with the data and without the establishment of restrictions or structural assumptions:

$$h_{it} = \frac{X_{it}}{N^{-1}\Sigma_{i=1}^{N}X_{it}} = \frac{b_{it}}{N^{-1}\Sigma_{i=1}^{N}b_{it}}$$

With the application of the previous formula, the authors manage to trace an individual trajectory for each object without depending on the common tendency of all the objects. Thus, it was possible to observe the exact trend of a certain object within the global trend μ_t.

Additionally for this transition path, in case there is convergence, there must be a common limit for each case studied. In this way, the hit coefficient will tend to unity for each object as time progresses ($h_{it} \rightarrow 1$ for all objects i = 1, 2,…,N when t→∞).

Also, the mean squared distance for the panel from the common boundary, the cross-sectional variation H_{it}, must converge to 0, so that:

$$H_t = N^{-1}\Sigma_{i=1}^{N}(h_{it} - 1)^2 \rightarrow 0 \quad when\, t \rightarrow \infty$$

In turn, to build a statistical test for convergence, Phillips & Sul (2007) [19] assume the following parametric estimation of b_{it}:

$$b_{it} = b_i + \frac{\sigma_i \xi_{it}}{L(t)t^{\propto}},$$

With b_i as a fixed value, invariant in time, the parameter ξ_{it}, as identically distributed independent random variables of N (0, 1) along *i* but dependent on *t*, L(t) is a slowly varying function in the time that approaches infinity as time approaches infinity (L(t)→∞ when t→∞), and $\propto$ the convergence ratio.

The hypothesis test for convergence would be:

$$H_0: b_i = b \text{ y } \propto \geq 0$$

$$H_1: b_i \neq b \text{ y } \propto < 0$$

If the null hypothesis is not ruled out, there may be different trajectories for the objects, including divergence.

Additionally, Phillips and Sul (2007) [19] and Phillips and Sul (2009) [20] propose to study the existence of convergence between objects by estimating the following model applying the method of ordinary least squares:

$$log\frac{H_1}{H_t} = -2\log(\log t) = \propto \; + \; \beta \log t + u_t, \quad para\, t = [rT], [rT]+1, \dots, T.$$

In this equation, $H_t = N^{-1}\Sigma_{i=1}^{N}(h_{it} - 1)^2$, y H_1/H_t is the variance ratio of the cross section; β represents the speed of convergence of the b_i parameter; -2log (log t) is a penalty function that improves the performance of the test under the alternative hypothesis; r has a positive value on the interval (0,1] to discard the first observation block from the estimation, finally, [rT] is the enter part of rT. Phillips and Sul suggest using r ∈(0.2, 0.3) for samples of small size (T < 50) as a result of Monte Carlo simulations.

The null hypothesis of convergence is contrasted by applying the one-tailed t-test at a significance level of 5%, rejected at this level if $t_{\beta <} - 1.65$. The test is robust to heteroscedasticity and autocorrelation (HAC) to inequality α > 0 (using the estimate β=2∝) The log-t test served to contrast the convergence of clubs for the entire sample. If this test is rejected, Phillips and Sul (2007) [19] and Phillips and Sul (2009) [20] propose repeating the procedure by applying a four-step grouping, described in the next section.

If the log-t test is rejected for the entire sample, we repeat the test with the following methodology:

1. We order the units in descending order according to the last observation of the period.

2. We formed a central group by running the log-t regression for the first K units, where 2 < k < N, maximizing this with the condition that t > -1.65. We can establish the size of the central group, k*, as follows:

$$k^* = \arg\max[t_k] \;\; s.a \; \min[t_k] \succ 1.65$$

In case $t_k \succ 1.65$ is not fulfilled for the first two units, k=2, we eliminated the first unit and repeat the procedure. If it happens that $t_k \succ 1.65$ does not hold for any of the chosen units, the entire sample diverges.

3. Selection of data belonging to the club. After detecting the group k^* we will run the log-t regression to add one by one the objects that do not belong to the central group. If est_k is greater than the critical value c^* this unit will be added to the convergence club.

All these units, including those belonging to the central club, will form the first convergence club.

4. Rule of repetition of the process and stop. If there are objects for which the previous condition fails, all of these must be grouped together and the log-t test must be executed again, to check if the condition $t_k \succ 1.65$is fulfilled in them. If it is fulfilled, we will have a new convergence club. If this is not the case, the previous steps will be repeated to identify possible subgroups that form convergence groups. If the condition $t_k \succ 1.65$is not fulfilled for these objects in step 2, we will conclude that they diverge. The authors insist that the condition $t_k \succ 1.65$must be fulfilled for the clubs. If this is not satisfied by the clubs, we can increase the value of c^* until it is satisfied.

Given that the number of clubs that the algorithm will be able to identify is subject to the constitution of a first central group and that the following derive from the position of this, the critical value that we set when we refer to C* = must be previously meditated. This value lies in the level and type of error that we are willing to take. So:

A high level of c* will be useful if we don't want to make mistakes when classifying objects in clubs to which they probably don't belong. Setting a high level of C* will lead to the formation of more groups, even being able to appear more than there really are.

A low level of c* will be useful if we pursue the goal of having fewer clubs and thus have greater interpretability and less complexity in defining groups.

Phillips and Sul (2007) [19] and Phillips and Sul (2009) [20] recommend that, in small samples ($T<50$), a $C^*=0$ be used.

Regarding the number of resulting clubs and as a solution to this problem, Phillips and Sul develop an algorithm for merging the clusters formed with the previous algorithm. This will be able to detect clubs close to each other by running the log-t test comparing the clubs one by one with the rest of the clubs.

Following the same criteria as above, if the t statistic takes a value less than –1.65, these clubs will not converge, if, on the other hand, the statistic is greater than –1.65, the clubs will converge and form a new club.

Figures

Figure 1: Convergence transition paths and speed in cigarette brands (in packs), 2005–2021.

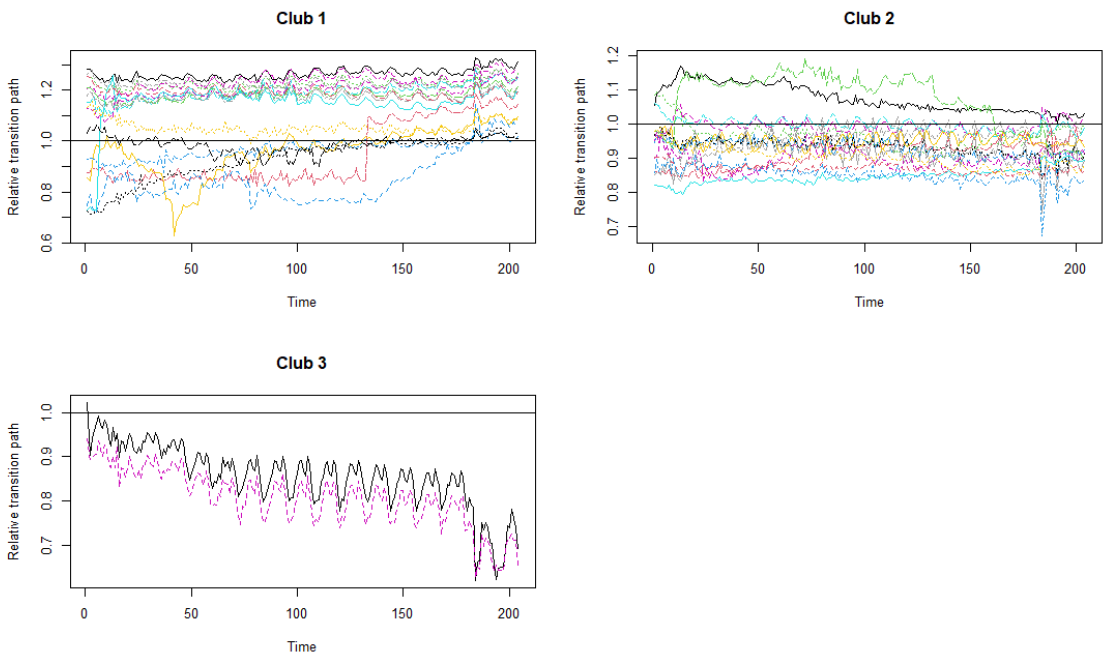

Figure 2: Convergence transition paths and speed in cigarette brands (in packs), 2005–2010.

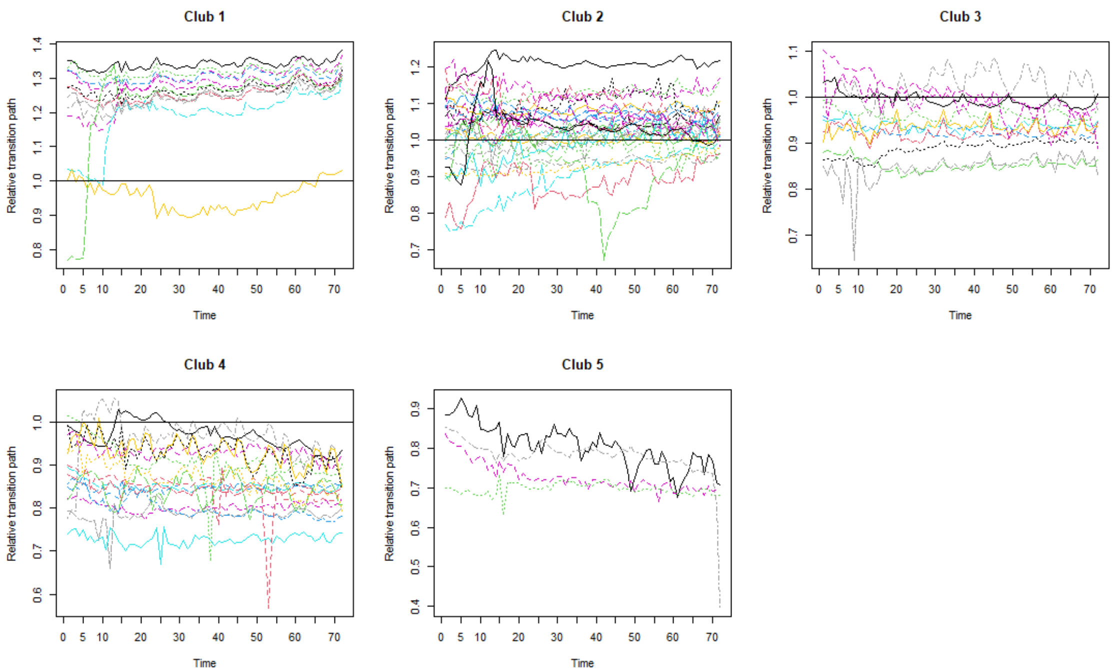

Figure 3: Convergence transition paths and speed in cigarette brands (in packs), 2011–2021.