Department of Economics and Finance

University of Guelph

Discussion Paper 2016-04

Detecting Convergence Clubs

By:

Fuat C. Beylunioglu, Thanasis Stengos, and M. Ege Yazgan

Detecting Convergence Clubs*

Fuat C. Beylunioglu† Thanasis Stengos‡ M. Ege Yazgan§

May 2016

Abstract

The convergence hypothesis, which is developed in the context of growth economics, asserts that the income differences across countries are transitory, and developing countries will eventually attain the level of income of developed ones. On the other hand convergence clubs hypothesis claim that the convergence can only be realized across groups of countries that share some common characteristics.

In this study, we propose a new method to find convergence clubs that combine pairwise method of testing convergence with maximal clique algorithm. Unlike many of those already developed in the literature, this new method aims to find convergence clubs endogenously without depending on priori classifications. In a Monte Carlo simulation study, the success of the method in finding convergence clubs, is compared with a similar algorithm. Simulation results indicated that the proposed method perform better than the compared algorithm in most cases. In addition to the Monte Carlo, a new empirical evidence on the existence of convergence clubs is presented in the context of real data applications.

Keywords: Growth Economics, Convergence Hypothesis, Convergence Clubs, Maximal Clique Algorithm.

JEL Classification: C32, O47.

*This work has been produced as a part of the research project (project no: 113K757) supported by The Scientific and Technological Research Council of Turkey (TUBİTAK). Ege Yazgan and Thanasis Stengos would like to acknowledge financial support from TUBİTAK.

†Istanbul Bilgi University, email: fuadcan@gmail.com

‡University of Guelph, email: tstengos@uoguelph.ca

§Istanbul Bilgi University, email: ege.yazgan@bilgi.edu.tr

1 Introduction

One of the main predictions of (neoclassical) economic growth theory is that in the long run, all countries with similar technological characteristics would converge to a balanced growth path (steady state) equilibrium that will be entirely determined by the (exogenously) given growth rate of technical progress, which in turn would equal labour productivity growth. Hence economies with the same productivity would grow at the same rate and converge to the same equilibrium. This is the so called growth convergence hypothesis, which has been one of the main focal points of the empirical economic growth literature. In that context, a time series interpretation of the convergence hypothesis considers income gaps (or labour productivity gaps) between countries over time and analyzes whether these gaps would diminish, hence signifying convergence to a single steady state (equilibrium). On the other hand, if there are constant or increasing returns to capital, there may be a multiplicity of steady states (or absence of stable steady states) and a country's initial conditions will determine to which of these it will converge. In essence, convergence to a single steady state implies that however poor, a country will inevitably converge in the long run to a (prosperous) equilibrium shared by all. In the absence of such a single steady state, poor countries may only converge to a common equilibrium with other poor countries and will never catch up with the prosperous ones. The current debate on growth convergence as it has evolved over the last three decades, has been one of the most active research areas in economics and has taken the central role in the empirical growth literature. The main developments in the literature as they have evolved over time, mainly since the mid eighties and are summarized and presented in the Durlauf et al. (2005) survey. The earlier literature was based on the analysis of standard cross section/panel data and it main contributions was to identify all the main issues that have arisen in that context such as such as endogeneity, heterogeneity and nonlinearity. However, lately the emphasis has been on utilizing the existing data sets of long time series of GDP data compiled for most countries after WWII (and for a fewer developed countries going back to the nineteenth century). The resulting time series approach has built on the work of Bernard and Durlauf (1995, 1996) who have introduced time series interpretations of the convergence hypothesis that can be cast in terms of unit root and cointegration analysis.

In that strand of recent literature, Pesaran (2007) has extended the time series convergence concepts to the case where there is no requirement that the converging economies to be identical in all aspects including initial endowments. The main result is that for two economies to be convergent it is necessary that their output gap is stationary with a constant mean, irrespective of whether the individual country's output is trend stationary and/or contains unit root. Furthermore, testing for convergence in that case does not rely on using a benchmark country in order to define the output gaps that are used in the analysis and uses a pair-wise approach to test convergence. The issue of relying on a benchmark, also renders the analysis problematic as perceived leaders used as benchmark economies may not retain the leader title over the whole period of analysis. In that respect, Pesaran's (2007) pair-wise analysis becomes relevant. This analysis only considers the binary process of convergence (or lack of it) for all pairs out of a set of countries included in the initial group. The choice of this initial group is arbitrary and usually accomplished based on the data availability, geographic or economic developmental status. Therefore, the analysis has nothing to say how if one can also examine the issue of convergence to a common cluster that can selected out of the initial group. Pesaran stated that "in principle, the convergence results from the analysis of pair-wise output gaps can be used to form "convergence clubs", but special care must be taken in addressing the specification search bias that such a strategy would entail" (Pesaran, 2007, p. 314). In other words the analysis so far, has mainly analyzed the issue of convergence between "country-pairs", but is mainly silent on how to proceed to classify countries as belonging to a common "country club". This is the main question that we examine in the current paper.

In this paper we try to examine the issue of different country groups converging to multiple steady states and the emergence of "convergence clubs" as was put forward by various researchers in the literature, see Baumol (1986), Durlauf and Johnson (1995) and Galor (1996) to name a few. We will proceed by introducing a new method that combines unit root testing within a $I(1)/I(0)$ framework with an the maximum clique approach from the computer science graph theory to establish a set of statistical criteria for cluster formation. We will also offer an evaluation of the performance of our proposed method vis-a-vis other existing methods in the literature by means of a Monte Carlo simulation. To the best of our knowledge, this is the first time that the properties of these methods have been explored and analyzed in the literature. The paper is organized as follows. In the next section we will discuss the relevant literature on club formation. We will then proceed to discuss in detail the competing approaches that we will be investigating in section 3 and then we will present the description of the Monte Carlo design and discuss the data generating processes, evaluation procedures and Monte Carlo results in Section 4. In Section 5 we will present the empirical results of the illustration of the method on real growth data and finally we will conclude.

2 Literature Review

The definition of a convergence club and the principle of clustering behind its formation gave rise to different empirical strategies to test the convergence hypothesis. However, the existing early methods were generally, focused on the convergence of various *a-priori* defined homogeneous country groups which were assumed to share the same initial conditions.Baumol (1986) for example grouped countries with respect to political regimes (OECD membership, command economies and middle income countries), Chatterji (1992) allowed for clustering that based on initial income per capita levels and tested convergence cross-sectionally, while Durlauf and Johnson (1995) grouped countries using a regression tree method based on different variables such as initial income levels and literacy rates that determined the different "nodes" of the regression tree that defined the country clubs with the common initial conditions and literacy characteristics. An alternative approach to the cross-sectional notion of $\beta-$ convergence in the context of cross-sectional was introduced by Bernard and Durlauf (1995, 1996) based on a time series framework that makes use of unit root and cointegration analysis, see Durlauf et al. (2005) for a comprehensive literature review for convergence hypothesis. Hausmann et al. (2005), similar to previous studies, by considering a priori grouping criteria such as initial incomes, found some evidence on convergence clubs by using time series methods.

In a time series context, Pesaran (2007) proposed a testing procedure that applies unit root tests to pairwise differences of the income per capita time series. This method relies on the use of unit root tests to all possible pairwise differences of the per capita income series in any given group of countries. Pesaran also considered different initial set of countries based on geographic characteristics for his pairwise method, but found no evidence on convergence clubs.

Similar to Durlauf and Johnson (1995), Hobijn and Franses (2000) (henceforth HF) proposed a panel data based approach for testing convergence. Contrary to the early attempts that relied on a two stage method that first assigns membership to a group and then considers whether this assignment is satisfied by the data, HF classifies countries into clusters of countries if they satisfy some criterion (desired convergence property). They clustered countries into subgroups by applying multivariate stationarity tests to panels consisting of pairwise differences of income per capita series and in contrast to Durlauf and Johnson (1995) they detected a larger number of small clubs. A different approach was proposed by Kapetanios (2003, 2008) who developed a method that is designed to endogenously classify stationary and nonstationary series by sequentially reducing the size of the null by removing series with the most evidence against the unit root null, classifying these series as stationary. The stopping point is when the unit root null does not reject, such that all the remaining regions are declared nonstationary.

Using the HF methodology, Corrado et al. (2005) extended this method by allowing subgroups to vary over time and applied it to European regional sectoral data of agriculture, manufacturing and services. Corrado and Weeks (2011) extended the sequential HF approach to account for short time panels by using a bootstrapping modification and applied their method to study regional European convergence. The main contribution of HF is that it does not require an a a-priori classification of country groups and detects group formation in an endogenous manner. A similar approach is advanced using the notion of $\sigma-$convergence by Phillips and Sul (2007) who developed an algorithm based on a log-t regression approach that clusters countries with a common unobserved factor in their variance. In the convergence literature, $\sigma-$ convergence as opposed to $\beta-$ convergence deals with the reduction in the variance of the cross country income distribution over time, see Quah (1996).

Following Pesaran (2007) and his testing procedure that applies unit root tests to pairwise differences of the income per capita time series convergence is reached when the proportion of rejections obtained from the pairwise unit root tests is greater than a certain threshold. He applied this method to country groups belonging to different geographical regions and found no evidence of convergence clubs. However, as is in most of the earlier studies, he subjectively a a-priori defined the country groups under consideration and did not propose an endogenous clustering method. The current paper aims at developing a convergence analysis technique of cluster (club) formation that relies on pairwise testing both in the simpler $I(0)$ or $I(1)$ framework as in Pesaran (2007) combined with the maximal clique algorithm widely used in graph theory from the computer science literature, see Bron and Kerbosch (1973) and Konc and Janezic (2007). Rather than testing a-priory grouped country clusters, the method explores all convergent groups in a list of N countries that was previously subjected to pairwise convergence tests within a $I(0)/I(1)$ or a long memory framework. Within a long memory framework this method has been introduced recently by Ozkan et al. (2014) on a limited scale to study club formation among small (exogenously) defined groups of homogeneous countries. We propose to use this approach as a new endogenous cluster formation method for the all available countries and not simply small predefined groups and analyze and compare its properties with the existing endogenous cluster formation mechanisms of HF since they both rely on testing the time series properties of the mean function of output gaps as opposed to the variance ($\sigma-$ convergence of Phillips and Sul (2007).[1] In our paper we will compare these two approaches, by means of an extensive Monte Carlo simulation study using evaluation criteria from the forecasting literature. This will be the first time that the properties of such mechanisms will be investigated and compared.

3 Methodology

The simple pairwise method and HF are both seeking convergence by searching *similarities* in movements of outcomes in the process of time. To this end both methods expect all pairs in a club to move around zero or a constant, in particular stationarity in difference of pairs. However there are several differences in approaches as well as the treatments of pairs. First, HF constructs clubs endogenously via a clustering algorithm that runs recursive stationarity tests. On the contrary, the pairwise method does not construct clubs, but tests the lists of clubs that are given exogenously. However our approach will combine pairwise testing with the maximal clique algorithm from computer science graph theory introduced by Ozkan et al. (2014). As it will be argued below, there is a crucial theme in the construction of a single club, HF is a bottom up method that forms the clubs by adding countries one by one while the maximal clique method, by definition of

[1] In the last few years, there are a number of papers that have looked at (club) cluster formation in different research areas. Abbott and De Vita (2013), Fritsche and Kuzin (2011), Abbott et al. (2012), Kim and Rous (2012), Apergis and Padhi (2013), Yilmazkuday (2013) and Ikeno (2014) to name a few.

employing the pairwise method, is a top down method that finds all the set of countries satisfying the definition of a club.

We proceed to present our proposed new convergence analysis technique that consists of pairwise testing as in Pesaran (2007) combined with a maximal clique algorithm widely used in graph theory from computer science literature. We first present the pairwise testing method and then the procedure to find convergence clubs via the maximal clique technique. We will proceed to compare our proposed method with that of HF by means of an extensive Monte Carlo simulation study.

3.1 Pairwise Convergence Test

Suppose that the log GDP per capita series of country i and j at time t are as follows

$$Z_t^{ij} = y_t^i - y_t^j = \beta + \varepsilon_t \quad \sim I(d),\ i = 1, ..., N,\ i \neq j,\ t = 1, ..., T$$

where T is the length of time interval, N is the number of countries and y_t^i and y_t^j denotes the log GDP per capita series of i and j. ε_t stands for the disturbance term and $d \in \{0, 1\}$ represents the integration of the series. Here β can represent a constant or a function of time as well. (see Stengos and Yazgan (2014)). Since the difference series are either stationary or non-stationary, that will determine if the pair is convergent or not. For instance if $\varepsilon_t \sim I(0)$, the two log GDP per capita series will be drifting together overtime and in that case it is appropriate to assert that countries i and j are convergent. On the other hand, if $\varepsilon_t \sim I(1)$, a nonstationary process would indicate that the log difference series between i and j is nonstationary and the two log GDP per capita series would be drifting apart over time, indicating that countries i and j are not converging.

Determining convergence by applying unit root tests on differences between GDP per capita series characterizes the time series based approach on convergence applied in many different contexts since Bernard and Durlauf (1995, 1996), see Durlauf et al (2005) for a comprehensive survey. However, when there are more than two countries, there is still uncertainty in determining whether the countries are converging altogether to a steady state. In the literature, the main approach centers on testing if all countries in the group are converging to the group average or a chosen country as a benchmark (generally United States), hence applying unit root tests to the pairwise differences of each group member with the average or the selected benchmark country. Alternatively, another approach is to apply multivariate stationarity tests to determine convergence. The former approach is criticized for the arbitrariness in choosing the benchmark country or the country average, while the latter is not preferred because of the difficulties in applying it to large groups.

The pairwise method developed by Pesaran (2007) can offer a remedy to both of the above difficulties. According to this approach, if one tests for convergence of a group of N countries, all $N(N-1)/2$ pairs are subjected to unit root testing. Pesaran (2007) showed that, if a group of N countries are non-convergent, the rejection rate of the null hypothesis of non-stationarity ($H_0 : Z_t \sim I(1)$) calculated by $N(N-1)/2$ tests is equal to the nominal size of the individual tests, i.e. the probability of Type 1 error. More specifically, it is shown that under the null hypothesis of N countries being non-convergent, the rejection rate of individual tests converges to the nominal size, α, as N and $T \to \infty$, even though individual tests are not independent cross- sectionally. Since the null hypothesis in this case is non-convergence (divergence) of N countries, in order to find evidence in favor of the null, it is enough to show that the proportion of rejections over $N(N-1)/2$ tests is not larger than the significance level of individual tests. In that case for example, if the significance level is 5%, the proportion of rejections must not exceed 0.05[2]. To summarize, rejection rates of $H_0 : Z_t \sim I(1)$, higher than a given significance level in a given application would imply evidence against the non-convergence (divergence) null hypothesis in favor of the convergence alternative. On the other hand, rejection rates lower or close to the employed significance level will provide evidence for the non-rejection (validity) of divergence.

3.2 Maximal Clique Method for Finding Convergence Clubs

The maximal clique method that we present in this subsection, combines the maximal clique algorithm of graph theory with the previously described pairwise convergence tests of $H_0 : Z_t \sim I(1)$. Rather than testing a priory grouped country clusters, the method explores all convergent

[2]No doubt, nominal size of the tests may differ from the significance level. In applications, the power of the tests used relative to size distortions should be given attention. Another matter to be attentive to is the fact that the rejection rate would converge to α in the limit and that of course would not be the case if N and $T$are relatively small in a given application.

groups in a list of N countries that was previously subjected to pairwise convergence tests. In this sense, the method is an endogenous extension of Pesaran (2007) similar to the one by HF.

The method consists of two steps. First, all possible pairwise differences of N countries are subjected to unit root tests where the null denotes a unit root process as evidence of non-convergence. If the rejection rate obtained from $N(N-1)/2$ tests is well above the significance level, that would be evidence against the null hypothesis of non-convergence (divergence) hypothesis in favor of the alternative of convergence and the list of N countries will be taken to form a convergent group. If this club involves all examined countries, then all countries are said to be convergent and we do not go any further in seeking out the presence of convergence clubs. However, as shown in Pesaran (2007), Dufrénot et al. (2012) and Stengos and Yazgan (2014) it is very unlikely that, by examining all countries as a single group, one will find evidence of convergence for all with pairwise testing. Nevertheless, if a subgroup of countries is found convergent via pairwise method, then it can be said that this subgroup constitutes a convergence club. The main challenge, as indicated above, is to find a method to determine this subgroup rather than relying on a-priori classifications. In the second step we undertake this task.

Assume that $\mathcal{U}$ denotes the set of all countries. Hence, by definition, the cardinality of $\mathcal{U}$ is equal to N; mathematically if $\#()$ denotes the cardinality, we have $\#(\mathcal{U}) = N$. Moreover, suppose that $\mathcal{E}$ is a subset of $\mathcal{U}$. In this case, in order $\mathcal{E}$ to be a convergence club, all binary combinations obtained with elements of $\mathcal{E}$ should satisfy the pairwise stationarity tests **(reject** $H_0 : Z_t \sim I(1)$**)** . Hence, since $\#(\mathcal{E}) = M < N$, the rejection rate obtained via all $M(M-1)/2$ pairs should be well above the significance level.

In the second step, from the $N(N-1)/2$ test results, the objective is to find a class of subsets $\mathcal{G}$ for which all subsets, e.g. $\mathcal{E}$, satisfy pairwise convergence property. Mathematically, let $\mathcal{G}$ denotes the class of all subsets satisfying the desired pairwise (stationarity) property. Then the problem is

$$\mathcal{G} := \{\mathcal{E} : \vee i, j, \ \ i \neq j, \ \in \mathcal{E}, \ t(Z^{ij}) = 1\}$$

where $Z^{ij} = y^i - y^j$, $t(\cdot)$ is the test result of the series in the bracelet and takes the value of 1 for a convergent pair, i, j and 0 otherwise.[3] Hence, the problem can be expressed as

[3]Notice that in order to satisfy the property explained above, for $\mathcal{E}$ to be a convergence group, all pairs $i, j \in \mathcal{E}, \ i \neq j$ should satisfy the convergence property. Conversely, we require the non-rejection rate of $H_0 : Z_t \sim I(1)$, which denotes divergence to be zero, a much more stringent condition, because it does not allow Type 1 error.

$$\arg\max_{\mathcal{G}}\{\ \#(\mathcal{E}):\ \ \mathcal{E}\in\mathcal{G}\}.$$

In graph theory terms, countries become vertices, the test result (rejecting or not rejecting pairs) of a pair become edges, and as such the set of all vertices and edges constitutes an undirected graph. If an undirected graph has edges between all vertices then the graph is said to be complete. If there is a subset of an undirected graph having all properties of a complete graph, the subset is so called a *clique*. Therefore, in our case, all convergence clubs of a country list can be expressed as cliques. Solving the problem defined above is known as finding maximal cliques.

Pairwise test results form an undirected graph and accordingly, countries and test results determine the vertices and edges respectively. Hence, the problem becomes to find a subgraph that has edges between each vertices, or in other words, a maximal clique. Figure 1 and 2 presents the notions mentioned above.

Finding a maximal clique can be too hard from a computational point of view. The computational complexity of solution to maximal clique problem is known as NP-Complete whose brute-force solution requires $2^N - \binom{N}{2} - N - 1$ trials. First, Bron and Kerbosch (1973) developed an algorithm to solve the problem in exponential time. In the recent literature, various planar graph algorithms have been developed that enables the problem to be solved in polynomial time. In this study, we will employ the branch and bound algorithm proposed by Konc and Janezic (2007) which is an improved branch-and-bound algorithm that ends in polynomial time.

We should note that, the maximal clique method is not a conclusive technique. In other words, it does not cluster the country list into subgroups, but finds club(s) having a maximum number of elements. Hence we offer the following clustering algorithm to detect convergence clubs.

1. Apply the desired stationarity test to all Z^{ij} such that $i, j \in \mathcal{U}$ and $i \neq j$.

2. Test the unit root null hypothesis. The resulting variable takes the value of 0 if null of non-stationarity (unit root) is not rejected, 1 if it is rejected (evidence for stationarity).

Similarly, we expect the rejection of $H_0 : Z_t \sim I(1))$ to be unity if convergence holds. One can relax this condition by allowing rejection rate up to a given level. It is worth noting that allowing the rejection rate up to a given level (say at the nominal size of a significance test) may have an impact on how much Type 1 error created by the unit root test would carry over in club formation. However, since our main goal is to compare different club formations mechanisms with Monte Carlo simulations relying on the same unit root tests, all these methods under comparison will be on the same footing. It could be argued that since in general unit root tests suffer from low power (where typically nominal sizes are lower than asymptotic significance levels) using a rejection rate close to zero can be interpreted as a precautionary step to confront the "low power" issue.

3. Construct adjacency matrix from the resulting variable values obtained in (2).

4. Find maximal clique(s) from the adjacency matrix via the algorithm proposed by Konc and Janezic (2007). If more than one clique is detected, proceed to the next step. If only one clique is detected jump to (6)

5. Select one of the detected cliques randomly and proceed to the next step.

6. The group of countries in the clique is labeled as a convergence club. Eliminate respective rows and columns of the countries from the adjacency matrix. And step back to (5). Stop if all the rows and columns are eliminated from adjacency matrix.

3.3 The HF Method

As mentioned in the introduction, the method presented above bears certain similarities to the endogenous cluster analysis proposed by HF. Hence, it is important to compare the accuracy of the our method with HF. To this end, we will first present HF and review both method by means of a simulation comparison in the following two subsections[4]. An important difference between the maximal clique method described above and the HF approach is that the former is based on testing the null of a unit root (divergence), whereas in the latter the null hypothesis is stationarity (convergence).

HF is a clustering algorithm that applies multivariate KPSS tests recursively to panels enlarged by a series in each iteration. More generally, the algorithm allows a new country to enter the convergence group until null hypothesis of stationarity is rejected.[5] HF relies on two definitions of convergence clubs. The first of these, perfect convergence, requires club members to have statistically equal GDP per capita series. Perfect convergence occurs if the pairwise difference of the club members' output series are stationary around a zero mean. This definition of convergence indicates a more stringent state, since it ignores catching-up possibilities or other differences stemming from initial conditions. The second definition of convergence, the so called relative

[4]As remarked in the literature review, other than HF, another method developed by Phillips and Sul (2007) stands out by means of not requiring a priori classification of countries. However, we exclude this method for the following reason. Unlike HF and our proposed pairwise maximum clique method, Phillips and Sul (2007) is based on σ type convergence. The method depends on the definition of convergence by means of reduction of variance over time and thus convergence of series to a steady state. Therefore, it is not appropriate to compare this method with HF and the method developed in this study as both of the latter deal with convergence of the mean (function) of the series.

[5]As is known, unlike ADF type unit root tests, KPSS tests directly the null hypothesis of stationarity.

convergence, describes similar movements in the output series over time irrespective of initial conditions, i.e. the pairwise differences follow the non-zero mean stationarity property. In other words, the expectation of the pairwise differences satisfies that $\mathbb{E}[y_t^i - y_t^j] = 0$.

As mentioned above, HF determines convergence of a group by utilizing multivariate KPSS tests. The method is based on the construction of a panel containing pairwise differences of consecutive series, then it applies KPSS test to this panel. In this manner, to test if country group $C = \{c_{n_1}, c_{n_2}, \cdots, c_{n_p} : n_p < N\}$ is converging, the panel $\mathbf{x}_t^C \equiv \mathbf{M}_p \mathbf{y}_t^C$ is defined where $\mathbf{M}_p$ and $\mathbf{y}_t^C \in \mathbb{R}^p$ are as follows.

$$\underset{(p-1)\times p}{\mathbf{M}_p} = \begin{bmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & -1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -1 \\ & & & & & \end{bmatrix} \quad \text{and} \quad \mathbf{y}_t^C = \begin{bmatrix} y_{1t} \\ \vdots \\ \vdots \\ y_{kt} \end{bmatrix}.$$

Here, $\mathbf{x}_t^C$ denotes the matrix of consecutive differences of incomes, $Z_t^{(i-1)i}$, $\forall i \in \{n_1, \ldots, n_p\}$, and the stationarity test applied to $\mathbf{x}_t^C$ determines whether or not country group C constitutes a convergence club. HF tested perfect and relative convergence separately by employing two respective multivariate KPSS tests. Convergence clubs from N countries are clustered via the following algorithm.

HF Algorithm:

1. Initially, set each countries as a club by defining $k_i = \{i\}$ for $i = 1, \ldots, N$.

2. For each $i < j$, construct $\mathbf{y}_t^C$ where $C = k_i \cup k_j$. Through these matrices, apply the multivariate KPSS test to $\mathbf{x}_t^C$. If null hypothesis of stationarity is rejected for all i, j, reject convergence hypothesis and stop. If it is not rejected for any pair, i, j proceed to next step.

3. Choose i, j that is tested to have largest p-value from the KPSS test in the previous step. For $i < j$, redefine k_i as $k_i = k_i \cup k_j$ and set $k_j = \emptyset$. Step back to (2).

4. Label non-empty sets obtained as convergence clubs.

3.4 Comparison of Methods

HF is a method that relies on a "bottom up" algorithm that clusters groups one by one. On the contrary, the maximal clique method relies on a "top-down" process that detects all subsets satisfying club properties. Other than clustering, there is a substantial difference in testing convergence. To determine whether a set of countries is convergent, HF applies multivariate stationarity test to panels comprised of consecutive pairwise difference series set elements and confirms convergence if the null hypothesis of stationarity of the panel is not rejected. However, the panels do not include all possible pairwise differences but only differences of consecutive pairs. For example, if we want to test the convergence of countries 1,2,3 and 7, a panel consisting of Z^{12}, Z^{23} and Z^{37} is subjected to the test, and if stationarity cannot be rejected the panel is then augmented to include additional difference series. If then in this "augmented" panel the stationary null is rejected, then these four countries are said to be convergent. On the other hand, our proposed pairwise method depends on a different definition of clubs, so that for N^* countries to be convergent, we need to achieve rejection of the null of a unit root for all $N^*(N^*-1)/2$ pairs. Hence, in order for the list of countries in the previous example to form a convergence club,the rejection rate of $4(4-1)/2 = 6$ pairs from unit root tests should exceed some significance level.

3.5 Monte Carlo Structure

In this subsection, we will discuss the data generating processes that is used in our Monte Carlo study. We generated various types of data to conduct the evaluation of the clustering methods that we compare in order to determine factors and sources leading to success and failure. The data sets are classified in two groups. In the first group we include single club and many non-convergent pairs, while the ones in the second group include multiple clubs together with only some non-convergent pairs. In the following parts of this subsection, we will present the data generating processes and evaluation procedures employed in this study.

3.5.1 Data Generating Processes

The simulation assumes that the log GDP series is given as follows.

$$y_{it} = c_i + \gamma_i f_t + \epsilon_{it} \tag{1}$$

where $\epsilon_{it} \sim I(0)$ is the error term and f_t is the common factor which affects all countries the same way (such as technology). If we assume non-stationarity of the factor, a pair of countries can only be convergent if both countries utilize the factor likewise. This can be possible if the country specific constants, γ_i that measure that effect are equal. In other words, for the pair i and j, if $\gamma_i = \gamma_j$, f_t is canceled out and $y_{it} - y_{jt}$ becomes $c_i - c_j + \epsilon_{it} - \epsilon_{jt}$. In this case, since the error terms are assumed to be stationary, we have $c_i - c_t + \epsilon_{it} - \epsilon_{jt} \sim I(0)$ and the pair i and j would be convergent by definition. Likewise, for any subset of countries having equal γ_i terms, all pairwise difference series in this subset would be stationary and hence these countries would constitute a convergence club. On the other hand, the constants, c_i, are country specific and are generated once for all data sets.

The non-stationarity of f_t is modelled under an ARIMA process following below

$$f_t = f_{t-1} + v_t, \quad v_t = \rho_v v_{t-1} + e_t, \quad e_t \sim iid\ N(0, 1 - \rho_v^2)$$

where we allow $\rho_v = \{0.2, 0.6\}$ as separate cases. Besides, we also allow the error term of the log GDP series in equation (1) to have serial dependence, following the specification below,

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + v_{it}, \quad v_{it} \sim\ iid\, N(0, \sigma_{v_i}^2(1 - \rho_i^2)).$$

Above, the error terms, v_{it} are i.i.d. distributed Normal random variables. Here, the autoregressive coefficient ρ_i and $\sigma_{v_i}^2$ are country specific and invariant among the data sets. To be more precise, before proceeding to data generation, we generated the coefficients to have the following property.

$$\sigma_{vi}^2 \sim\ iid\, \mathcal{U}[0.5, 1.5], \quad \rho_i \sim\ iid\, \mathcal{U}[0.2, 0.6]$$

To generate a single club containing data set, the coefficients of m convergent countries, are assumed to be $\gamma_i = \gamma_j = 1$. For the remaining $(N - m)$ countries, γ_i generated randomly as $\gamma_i \sim iid\ \mathcal{X}_{\kappa_i}^2$. It is worth noting that γ_i are generated once, yet, when the number of club members (m) is 10 instead of 5 for example, arbitrarily selected 5 of the remaining coefficients are substituted with 1 to allow them to be convergent. Lastly, we also generate country specific constants as $c_i \sim iid\ \mathcal{X}^2_{\kappa_i}$.

For multiple clubs, however, we introduced some additional modifications. We ensured two non-convergent countries to exist in all data sets by omitting 1 country from the last two clubs. For example when $N = 20$ and $k = 4$; we let $m_1 = m_2 = 5$ and $m_3 = m_4 = 4$ where m_{k_n} denotes the number of elements of the k_n'th club. Furthermore, γ_k are equal for countries constituting a club, but unequal among clubs. In particular, γ_k are chosen from $\{1, 3, 4, 7, 8, 10\}$ and the remaining two coefficients of non-convergent countries are generated as being in single clubs.

The simulations are repeated 10000 times for various time intervals; number of countries, clubs and club members. In particular, simulations are performed using different combinations of $T = \{50, 75, 100\}$ time intervals, $N = \{10, 20, 30, 40\}$ count of countries and $k = \{1, 2, 3, 4, 5, 6\}$ number of clubs. When there is $k = 1$ convergence club, $m = \{5, 10\}$ of club members are considered and if $k > 1$, $m = N/k$ and $m = N/k - 1$.

3.5.2 Testing and Evaluating Procedures

We start by generating the repeated data sets based on the number of replications following the above specifications and choice of parameters. Both methods are applied to each generated data sets and the resulting club(s) obtained via both methods are evaluated by comparing the predicted club formation of each method with the actual club formation from the data generating processes described above. The general evaluation of the success of each method is considered for each data type.

To the best of our knowledge, since there is no other comparable Monte Carlo study in the literature that evaluates clustering methods in the same context as we do here. We will make use of some statistics from other fields to evaluate the maximal clique algorithm and HF and the methods that we will propose differ for single club to the multiple club cases. We will first present the single club evaluation methods, for which we will utilize three different cases. The first two, the Kupiers Score (KS) and the Pesaran and Timmermann (1992) Test Statistics (PT) are commonly used in the forecasting times series literature, while the third one, perfect detection, is a common evaluation method which is quite intuitive. Both PT and KS are used for evaluation of sign forecasts. It is worth noting that sign forecasts are used for predicting whether an underlying series would appreciate (increase) or depreciate (decrease) relative to a benchmark. In our case, success in detecting a country's membership in a club is equivalent to success in forecasting the sign of a time series.

Since success of bidirectional results such as upside and downside movement or membership in a club, can occur randomly[6], KS takes true forecasts and false alarms into account separately. For instance, if we are evaluating a forecast of a bad event or calamity in economics, an estimate of false alarms would help us avoid the issue of scare-mongering. Here, KS is defined as $H - F$ where

$$H = \frac{II}{II + IO}, \text{ and } F = \frac{OI}{OI + OO}$$

The capital letters "I" and "O" in the above formulae refers to whether the country under investigation is a member ("in" the club) or not ("out" of the club). Regarding the order of the letters, the first letter indicates whether the country is found to be a member in the experiment, while the second letter denotes its actual membership state (i.e. whether the country is actually in the club or not). Therefore, "II" indicates that the country, as a member of the club is correctly identified whereas, "OO" denotes that that the country, as an outsider of the club is also correctly identified. Furthermore, IO indicates that a country is detected to be in the club, while actually it is not (false detection). Similarly OI refers to the case where the country is mis-classified as being outside, even though it is a member of the club (false alarm). The ratio H captures the rate of "correct hits" in detecting club membership, whereas F denotes the "false alarm" rate, that is the rate of false exclusions.

As in the case of sign prediction in the forecasting literature, success can be the outcome of a pure chance probability event of 0.5. Hence, to test the statistical significance of KS, we will employ the following PT statistic,

$$PT = \frac{\widehat{P} - \widehat{P^*}}{[\widehat{V}(\widehat{P}) - \widehat{V}(\widehat{P^*})]^{\frac{1}{2}}} \sim N(0, 1),$$

where $\widehat{P}$ refers to the proportion of correct predictions (correct detections of countries as being a member or non member) over all predictions (N countries), and $\widehat{P^*}$ denotes the proportion of correct detections under the hypothesis that the detections and actual occurrences are indepen-

[6]This is similar to expecting an unbiased coin to come up heads with 50% probability.

dent.(where success is a random event of probability 0.5), while $\widehat{V}(\widehat{P})$ and $\widehat{V}(\widehat{P^*})$ stand for the variances of $\widehat{P}$ and $\widehat{P^*}$ respectively.

In simulations involving multiple clubs, it is not possible to use KS and the PT statistic due to the more complicated nature of the success/failure classification which is no longer binary as in the case of the single club case. In the situation of a single club, a country can be either detected (correctly or incorrectly) to be a member of this single club or not. In the multiple club case, on the other hand, a country can be correctly found to be a member of a club, but this club may be the wrong one (i.e. a club in which the country does not belong to in reality). In other words, there are more than two distinct cases for the actual membership state: the country can be either a member of the correct club, or belong to the "wrong" club, or not be a member of any club

To confront this problem, in the case of multiple clubs we will use a much stricter criterion by counting the successful cases in our simulations in which *all* countries are detected correctly to be in their correct positions. In other words, we do not evaluate success as a binary outcome, country by country as in the case of a single club in each replication. Instead, we will look at the overall results in each replication. If, in one replication, all countries are placed correctly in their correct position we will consider this as one successful outcome out of a total of 10000 replications. Since this more stringent indicator of success can also be applied to the single club case, we also report it in that case in addition to the KS and PT criteria.

3.6 Simulation Results

We now proceed to discuss the findings of the simulations based on the data generating processes of club formation discussed above. We will first discuss the comparison between the pair-wise unit root based approach augmented with teh maximum clique algorithm and the HF approach based on multivariate KPSS testing for the single club case and then the multiple club case.

3.6.1 Single Club Results

The results from the single club analysis are presented in tables 2 to 1. In each table we have two choices of the number of club members m (5 and 10), four choices of the total number of countries involved N, three choices of time span for the analysis that would mimic the real data time span availability and two choices of the persistence parameter ρ. We consider 3 different ways that the pair-wise unit root testing approach could be combined with the maximum clique algorithm, one that use an ADF test, the second using the ADF with a GLS corrected unit root test and a third one using the KPSS test. HF is based in the multivariate KPSS testing procedure. Among the three versions of the pair-wise approach, the ADF one gives better and more consistent results that overall outperform its all competitors including HF. The last set of columns for example in table 1, for the $H-F$ results (the "correct hit" ratio net of "false alarms", indicates that with $m=10$, $N=40, T=100$ and $\rho=0.6$, ADF with a 0.947 KS outperforms the others including the HF method that has a KS of 0.844. Similarly for the case from table 2 the PT test statistics yield 578.067 for the pair-wise ADF test versus 551.074 for HF with success percentages of 63.8% and 48.8% respectively. Note, that the rejections of the null hypothesis of random success outcomes are very strongly rejected with the PT test by both methods (slightly more so by the pair-wise ADF approach). Overall, the GLS and KPSS based pair-wise methods are not doing as well as HF and certainly much worse than the ADF version for the case where there is no constant in the data generating process. For that reason we will be conducting the pair-wise analysis in what follows based on ADF alone.

The results change with the presence of a constant. In that case all three deferent versions of the pair-wise approach outperform HF in nearly all settings of parameter combinations both for the KS and PT statistics. The presence of the country specific constants in equation 1 makes for a more realistic setting and in that case it is clear that the pair-wise approach irrespective of the choice of unit root test results in more accurate detections and a more reliable club (cluster) formation mechanism.

3.6.2 Multiple Club Results

The results for the multiple club case are presented in table 5. Recall that we allow in that case for $k=\{1,2,3,4,5,6\}$ number of clubs and the number of club members is given depending on the case as $m=N/k$ and $m=N/k-1$. We only present the ADF version of the pair-wise method as it was clear from the previous single club analysis that the other two versions were outperformed by the simple ADF. Again, the pair-wise ADF method outperforms HF in the majority of cases and especially when N (the pool of countries available) and T, the time span increases, both in the case of the presence of a constant or not. For example, with $N=20$, $m=5$, $T=100$ and

$\rho = 0.6$ the pair-wise ADF method detects 90.1% correct classifications without the constant and 79.8% cases with the constant DGP, while HF detects 57.7% and 34% such cases respectively. The results suggest that in terms of accuracy the ADF-maximum clique augmented pairwise method does quite well in detecting correctly the presence of clubs or clusters of countries. This gives us confidence that using the above method to real data would provide us with useful insights about how countries over time collect themselves into different groups and club formations of similar characteristic as far as economic activity is concerned.

4 Real Data Application: Growth Convergence

Using as guidance of the Monte Carlo results presented above, we now proceed in this section to apply the club formation methods analyzed earlier using the GDP per capita data from the Maddison Project Database.[7] We considered five different types of data according to different time-span and country classifications. The first two data sets are based on data availability starting from 1930 and 1940. The other three are selected by both data availability from 1950 and the inclusion in Europe, the Group of Seven (G7) and the S&P Emerging Markets classification. Table 4 displays the list of the countries covered under each classification.[8]

Table 5 displays the number clubs that are found by HF and maximum-clique algorithms using 5 percent significance level. Unlike, the previous Monte Carlo simulations, where the appropriate choice is adopted depending only for the DGP contains a constant term, we simultaneously use HF test relative convergence.[9] For example if we look at the results obtained at 5 % significance level for 1930 group ($N = 36$), pairwise-Max-Clique approach based on ADF tests finds 72 clubs with 2 countries (# 2), 44 clubs with 3 countries (# 3) and finally 7 clubs with 4 countries (# 4). It is worth noting that while there are no clubs with a number greater than 4 countries, HF indicates a single club with 5 countries. Recall that the search of these convergence clubs is done over the total of 315 ($= N(N-1)/2 = 36(35)/2$) country pairs. As explained above, the pairwise-MaxClique approach does not exclude the possibility of common countries in the clubs with the same number of countries. In that case for instance, at least some of 7 clubs with 4

[7] The Maddison-Project, http://www.ggdc.net/maddison/maddison-project/home.htm, 2013 version.

[8] Notice that "1930" and "1940" groups are identical in terms of the countries included, however, naturally differ in terms of data length.

[9] For the pairwise method the lags are selected automatically by the Akaike Information Criterion.

countries should be expected to contain the same countries. On the hand HF as a result of its algorithm categorizes the list of all countries as convergence clubs with distinct (non-overlapping) elements. Therefore, the 5 clubs with 4 countries contain necessarily distinct countries. Overall, however, the pairwise-Max-Clique approach based on ADF and HF provide similar results.

The results above should be interpreted as the number of possible clubs detected by the different approaches. In this sense the pairwise and HF methods are not strictly comparable due to their structure. In particular the pairwise method with its graph extension is a top-down process which initially finds all converging pairs and then constructs the cliques. Thus the above tables present all the possible clubs including all subsets of the largest groups. The first columns of the tables include all converging pairs, while the other columns show the cliques that can be constructed via the pairs (if there are any). On the other hand, HF is a bottom-up process that finds only the largest cliques. Therefore in order to make the comparison between these two methods more accurate, we augmented the counts of HF by adding all sub-clubs, $\binom{n}{k}$ where $2 \leq k < n$. For instance if a club with six elements is detected then we add $\binom{6}{5}$ elements to the fifth column, $\binom{6}{4}$ elements to the fourth column and so forth. This refinement is expected to help with the the interpretation of the results and its comparison with the pairwise method[10].

5 Conclusions

In this paper we have introduced a new method that combines unit root testing within a $I(1)/I(0)$ framework with the maximum clique approach of graph theory to establish a set of statistical criteria for cluster formation. We offer an evaluation of the performance of our proposed method vis-a-vis the HF method in the literature, that is closer in spirit to our approach, by means of a Monte Carlo simulation. To the best of our knowledge, this is the first time that the properties of these methods have been explored and analyzed in the literature. In the application we encountered *almost* the same patterns as in the single club simulations. The results of the HF tests do not differ when the significance level changes, while on the contrary the pairwise test results differ dramatically. Compared to other methods, KPSS used with the maximal clique extension shows

[10]There may be also differences due size distortions and the presence of structural breaks. For instance, the data starting from 1930 and 1940 include the same countries, however the data starting from 1930 cover World War II and the post war periods, something that certainly affects the results.

large over-forecasting tendencies that decrease as the significance level increases. This is the same pattern that we obtained in the simulations for $n = 30$, $d = 1$ and $T = 50, 100$ where the model generates false alarms and becomes more precise as the significance level increases.

For the ADF-GLS, however, the results are harder to interpret. Simulation results show that ADF-GLS increases its success as T increases from 50 to 100 and this improvement appears to stem from the bias scores as a reduction in false detections. In the applications we see that the number of converging pairs are higher for the 1940 data than 1930. The difference seems to stem from the nature of the data along with unit root type structure of ADF-GLS, since we see the same pattern for ADF but not KPSS or HF. On the other hand, the ADF-GLS results are more conservative for the regional data that cover 61 time steps, something that accounts for the low PT results and high tendency to make false detections for $T = 50$.

In the near future we plan to apply this methodology to other empirical environments such as different stock market indices to examine the question of market efficiency across countries. What we have done here is compare our proposed method that is a top down approach to the simple bottom up approach of HF. In future research we also plan to extend our analysis to also cover alternative variants of the HF approach as in Corrado and Weekes (2011) and the alternative method proposed by Kapetanios (2003, 2008). It is worth noting that the methods proposed and examined here are based on an analysis of the mean function and they do not account for $\sigma-$convergence as in Phillips and Sul (2007). Examining the properties of the latter method is also left for future research.

5.1 Tables and Figures

Table 1: Single Club: Kupiers Scores, PT and Success Averages, with constant, 5 % significance level

Data Type				H		F		KS		PT		Perfect Detection	
m	N	\rho	T	ADF	HF	ADF	HF	ADF	HF	ADF	HF	ADF	HF
5	10	0.2	50	0.909	0.810	0.061	0.297	0.848	0.513	266.024	163.140	53.3%	35.7%
			75	0.982	0.877	0.075	0.299	0.907	0.578	286.441	185.590	79.4%	46.4%
			100	0.988	0.909	0.091	0.299	0.898	0.610	283.092	197.151	77.6%	52.4%
		0.6	50	0.919	0.824	0.033	0.262	0.885	0.561	279.750	178.222	59.1%	41.2%
			75	0.996	0.886	0.031	0.283	0.964	0.603	303.802	193.542	91.6%	49.9%
			100	0.995	0.911	0.034	0.277	0.961	0.635	304.334	204.359	93.0%	55.4%
	20	0.2	50	0.848	0.677	0.078	0.216	0.770	0.460	337.965	189.406	44.7%	29.0%
			75	0.958	0.772	0.081	0.245	0.877	0.528	371.469	210.967	62.6%	39.8%
			100	0.974	0.831	0.094	0.245	0.880	0.586	369.383	232.441	59.4%	47.6%
		0.6	50	0.887	0.715	0.032	0.198	0.854	0.517	381.145	213.397	55.3%	35.3%
			75	0.977	0.805	0.034	0.222	0.943	0.583	411.849	233.894	85.5%	45.1%
			100	0.983	0.853	0.034	0.226	0.949	0.627	413.927	249.767	86.5%	52.6%
	30	0.2	50	0.849	0.580	0.097	0.194	0.752	0.386	368.905	180.220	4.7%	3.6%
			75	0.964	0.672	0.135	0.214	0.829	0.458	389.498	205.735	0.4%	6.4%
			100	0.983	0.724	0.120	0.222	0.863	0.502	393.213	222.420	0.1%	9.7%
		0.6	50	0.863	0.616	0.064	0.177	0.799	0.440	403.882	207.360	11.5%	7.6%
			75	0.968	0.696	0.094	0.197	0.874	0.499	423.053	226.959	4.4%	14.2%
			100	0.983	0.756	0.092	0.206	0.891	0.550	423.839	245.565	1.5%	20.9%
	40	0.2	50	0.740	0.522	0.091	0.161	0.649	0.360	352.316	186.311	4.1%	3.1%
			75	0.913	0.611	0.109	0.184	0.804	0.426	408.563	209.332	0.3%	5.5%
			100	0.945	0.666	0.110	0.192	0.835	0.475	420.613	229.063	0.1%	8.6%
		0.6	50	0.806	0.558	0.067	0.148	0.738	0.410	432.256	214.717	11.0%	7.6%
			75	0.950	0.646	0.067	0.175	0.883	0.471	475.254	232.583	4.2%	13.9%
			100	0.971	0.708	0.073	0.181	0.899	0.527	481.261	255.646	1.5%	20.2%
10	20	0.2	50	0.821	0.593	0.058	0.192	0.763	0.401	344.995	183.646	8.6%	7.9%
			75	0.981	0.696	0.073	0.225	0.908	0.471	406.572	211.398	63.8%	20.3%
			100	0.991	0.770	0.095	0.231	0.897	0.539	405.349	240.995	65.4%	31.4%
		0.6	50	0.829	0.611	0.022	0.172	0.806	0.439	363.830	201.263	9.8%	9.9%
			75	0.985	0.713	0.029	0.209	0.956	0.504	427.786	226.005	80.8%	23.6%
			100	0.996	0.780	0.032	0.223	0.965	0.557	432.276	249.279	89.4%	34.9%
	30	0.2	50	0.803	0.508	0.096	0.177	0.707	0.331	388.236	188.753	1.7%	0.4%
			75	0.972	0.596	0.145	0.200	0.828	0.396	439.032	217.001	1.2%	2.0%
			100	0.993	0.656	0.136	0.214	0.857	0.441	442.321	237.300	0.2%	3.8%
		0.6	50	0.807	0.521	0.064	0.161	0.743	0.360	411.742	206.931	3.4%	1.3%
			75	0.973	0.611	0.101	0.185	0.872	0.426	464.779	234.679	8.2%	4.7%
			100	0.991	0.679	0.106	0.198	0.885	0.482	465.682	259.379	3.0%	10.2%
	40	0.2	50	0.798	0.468	0.070	0.143	0.728	0.325	454.583	213.460	1.8%	0.5%
			75	0.971	0.551	0.103	0.172	0.867	0.380	504.682	235.178	1.2%	1.5%
			100	0.990	0.620	0.112	0.180	0.878	0.440	507.887	265.686	0.2%	3.3%
		0.6	50	0.805	0.493	0.055	0.131	0.750	0.362	487.600	238.662	3.5%	1.4%
			75	0.973	0.580	0.066	0.158	0.906	0.422	539.405	262.542	7.5%	4.8%
			100	0.990	0.646	0.076	0.170	0.914	0.476	541.570	287.462	2.7%	9.5%

Table 2: Single Club: Kupiers Scores, PT and Success Percentages, without constant, 5 % significance level

Data Type				H		F		KS		PT		Perfect Detection	
m	N	\rho	T	ADF	HF	ADF	HF	ADF	HF	ADF	HF	ADF	HF
5	10	0.2	50	0.997	0.905	0.022	0.022	0.975	0.883	306.038	280.085	93.2%	66.6%
			75	0.997	0.945	0.034	0.027	0.963	0.918	305.017	290.568	93.4%	79.3%
			100	0.999	0.960	0.030	0.027	0.968	0.933	304.955	295.175	93.2%	83.8%
		0.6	50	0.997	0.907	0.024	0.019	0.973	0.888	304.683	281.627	93.1%	67.3%
			75	0.999	0.946	0.025	0.022	0.974	0.924	305.436	292.306	94.3%	80.0%
			100	0.998	0.961	0.038	0.022	0.960	0.939	304.430	296.894	94.2%	84.6%
	20	0.2	50	0.989	0.894	0.027	0.016	0.962	0.878	423.768	400.526	90.7%	66.1%
			75	0.992	0.936	0.030	0.020	0.962	0.916	419.205	409.866	90.9%	78.9%
			100	0.992	0.952	0.037	0.024	0.955	0.928	415.879	411.933	89.5%	83.2%
		0.6	50	0.985	0.898	0.029	0.014	0.955	0.884	419.034	403.638	90.5%	66.9%
			75	0.991	0.939	0.027	0.019	0.964	0.920	418.665	412.163	91.4%	79.4%
			100	0.992	0.954	0.028	0.020	0.964	0.934	420.265	415.803	92.0%	83.9%
	30	0.2	50	0.979	0.855	0.040	0.021	0.938	0.834	479.708	464.974	62.1%	53.0%
			75	0.986	0.904	0.054	0.023	0.931	0.882	466.983	479.694	51.4%	65.5%
			100	0.987	0.925	0.058	0.024	0.929	0.901	459.706	485.234	43.0%	71.6%
		0.6	50	0.982	0.875	0.034	0.015	0.948	0.860	488.076	480.616	72.9%	59.2%
			75	0.984	0.918	0.042	0.019	0.942	0.900	477.588	490.677	65.9%	71.7%
			100	0.987	0.939	0.047	0.020	0.940	0.919	475.142	495.842	61.1%	77.6%
	40	0.2	50	0.978	0.851	0.030	0.015	0.948	0.836	550.813	538.965	63.6%	53.6%
			75	0.981	0.899	0.047	0.017	0.933	0.882	533.726	553.031	51.8%	66.5%
			100	0.981	0.916	0.055	0.022	0.927	0.894	517.268	549.864	43.2%	70.9%
		0.6	50	0.974	0.871	0.036	0.012	0.938	0.859	551.093	555.208	73.5%	59.7%
			75	0.984	0.916	0.035	0.014	0.949	0.902	542.319	567.346	66.9%	72.5%
			100	0.982	0.932	0.048	0.018	0.935	0.915	531.894	564.958	61.2%	77.2%
10	20	0.2	50	0.989	0.738	0.024	0.014	0.965	0.724	433.004	334.246	85.3%	14.0%
			75	0.998	0.834	0.029	0.019	0.969	0.815	433.895	368.393	92.1%	37.8%
			100	0.997	0.881	0.037	0.021	0.960	0.860	431.632	386.371	90.7%	53.4%
		0.6	50	0.989	0.740	0.025	0.012	0.963	0.728	430.890	335.892	85.5%	14.2%
			75	0.999	0.837	0.025	0.016	0.974	0.820	433.809	370.786	93.0%	38.3%
			100	0.999	0.883	0.026	0.018	0.973	0.865	434.515	388.601	93.7%	53.9%
	30	0.2	50	0.984	0.709	0.040	0.019	0.944	0.690	507.274	412.069	61.5%	10.7%
			75	0.995	0.803	0.058	0.022	0.936	0.781	500.727	448.123	55.1%	30.3%
			100	0.996	0.852	0.065	0.024	0.931	0.827	494.699	465.922	46.4%	44.6%
		0.6	50	0.985	0.722	0.032	0.014	0.953	0.708	514.192	422.094	70.9%	12.2%
			75	0.995	0.817	0.041	0.018	0.954	0.799	509.760	457.643	69.4%	33.6%
			100	0.996	0.865	0.049	0.019	0.947	0.846	506.653	475.984	64.0%	49.2%
	40	0.2	50	0.983	0.708	0.029	0.013	0.953	0.694	588.487	487.841	63.0%	10.8%
			75	0.993	0.805	0.048	0.016	0.945	0.789	578.987	527.651	55.8%	31.1%
			100	0.993	0.848	0.056	0.020	0.937	0.828	567.580	541.185	46.5%	44.3%
		0.6	50	0.983	0.720	0.035	0.011	0.948	0.710	588.492	497.555	71.2%	12.1%
			75	0.994	0.816	0.035	0.013	0.959	0.803	585.736	535.908	69.9%	34.3%
			100	0.995	0.861	0.048	0.017	0.947	0.844	578.067	551.074	63.8%	48.8%

Table 3: Multiple Clubs: Success Percentages, 5 % significance level

Data Type				No Constant		With Constant	
m	N	\rho	T	ADF	HF	ADF	HF
10	2	0.2	50	90.1%	56.3%	29.4%	29.3%
			75	93.5%	68.1%	80.7%	39.5%
			100	92.9%	74.6%	82.9%	46.5%
		0.6	50	88.1%	56.5%	31.8%	33.7%
			75	92.0%	68.6%	89.4%	43.1%
			100	92.4%	75.1%	93.8%	49.4%
	3	0.2	50	86.0%	62.4%	24.5%	27.3%
			75	87.4%	71.3%	44.4%	38.1%
			100	85.5%	76.2%	41.0%	45.5%
		0.6	50	86.2%	63.1%	35.7%	35.9%
			75	89.0%	72.1%	72.0%	44.5%
			100	88.8%	77.0%	73.1%	50.7%
20	4	0.2	50	82.8%	25.2%	5.7%	9.2%
			75	89.2%	42.8%	49.7%	18.7%
			100	87.3%	53.8%	49.3%	27.1%
		0.6	50	82.9%	25.5%	7.5%	13.2%
			75	90.7%	43.3%	74.4%	23.4%
			100	90.3%	54.3%	80.6%	31.8%
	5	0.2	50	83.2%	31.4%	8.4%	11.5%
			75	88.7%	47.5%	49.9%	21.0%
			100	87.2%	57.0%	49.2%	29.1%
		0.6	50	83.1%	32.2%	11.3%	16.4%
			75	89.9%	48.2%	74.6%	26.7%
			100	90.1%	57.4%	79.8%	34.0%

Data Type				No Constant		With Constant	
m	N	\rho	T	ADF	HF	ADF	HF
30	5	0.2	50	78.4%	7.8%	1.1%	3.3%
			75	89.2%	23.6%	49.4%	10.7%
			100	87.4%	37.1%	54.5%	18.2%
		0.6	50	77.5%	8.1%	1.4%	4.9%
			75	89.7%	24.1%	70.3%	14.0%
			100	89.6%	37.3%	83.3%	21.7%
	6	0.2	50	81.3%	11.9%	2.1%	4.7%
			75	88.4%	28.6%	49.7%	12.5%
			100	86.7%	41.0%	50.6%	20.1%
		0.6	50	80.6%	12.2%	2.8%	6.9%
			75	89.3%	29.0%	71.9%	16.1%
			100	89.1%	41.2%	81.4%	24.4%
40	7	0.2	50	73.6%	3.3%	0.3%	1.7%
			75	89.4%	14.3%	45.8%	6.7%
			100	87.9%	25.7%	54.8%	13.6%
		0.6	50	72.7%	3.4%	0.4%	2.4%
			75	90.0%	14.3%	64.4%	8.6%
			100	89.8%	26.1%	82.9%	16.0%
	8	0.2	50	78.7%	5.8%	0.7%	2.1%
			75	88.5%	19.2%	47.0%	8.7%
			100	87.1%	30.8%	51.2%	15.6%
		0.6	50	77.2%	5.9%	1.0%	3.5%
			75	89.6%	19.4%	68.3%	11.0%
			100	89.6%	31.1%	80.7%	18.5%

Table 4: Country Groups based on Economic Characteristics and Data Availability for Growth Application

1930 & 1940	Germany, USA, Argentina, Australia, Austria, Belgium, UK, Brazil, Denmark, Ecuador, Finland, France, Guatemala, South Africa, India, Netherlands, Ireland, Spain, Sweden, Switzerland, Italy, Japan, Canada, Colombia, Costa Rica, Mexico, Norway, Peru, Portugal, Sri Lanka, Chile, Turkey, Uruguay, Venezuela, New Zealand, Greece
Europe	Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, UK, Ireland, Greece, Portugal, Spain, Albania, Bulgaria, Hungary, Poland, Romania
G7	Canada, France, Germany, Italy, Japan, UK, USA
S&P	Brazil, Chile, Colombia, Mexico, Peru, Hungary, Poland, China, India, Philippines, Thailand, Taiwan, Malaysia, Turkey, Egypt, Fas, South Africa

Table 5: Application on Growth Convergence: 5% significance level

Data / T /n	Type	# 2	# 3	# 4	# 5	# 6
1930	ADF	72	44	7		
T=81,N=36	HF	31	14	5	1	
1940	ADF	97	106	55	16	2
T=71,N=36	HF	42	25	8	1	
Europe + G7	ADF	37	5			
T=61,N=26	HF	31	22	10	2	
Europe + S&P	ADF	38	14			
T=61,N=38	HF	35	14	2		
G7 + S&P	ADF	27	9			
T=61,N=26	HF	20	11	5	1	

Figure 1: A sample undirected graph

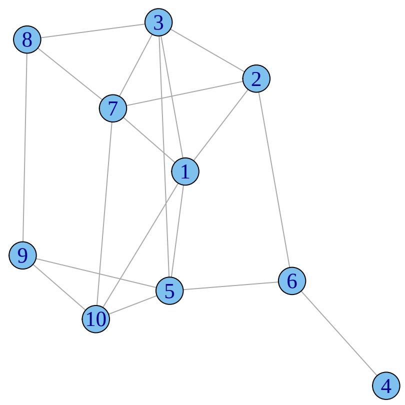
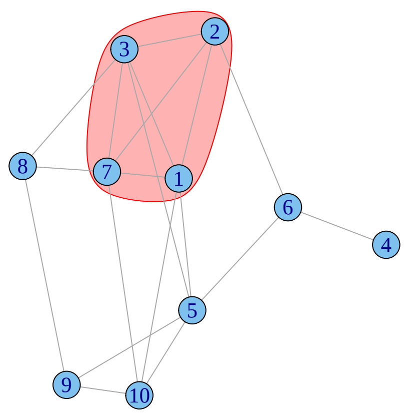


Figure 2: A sample maximum clique



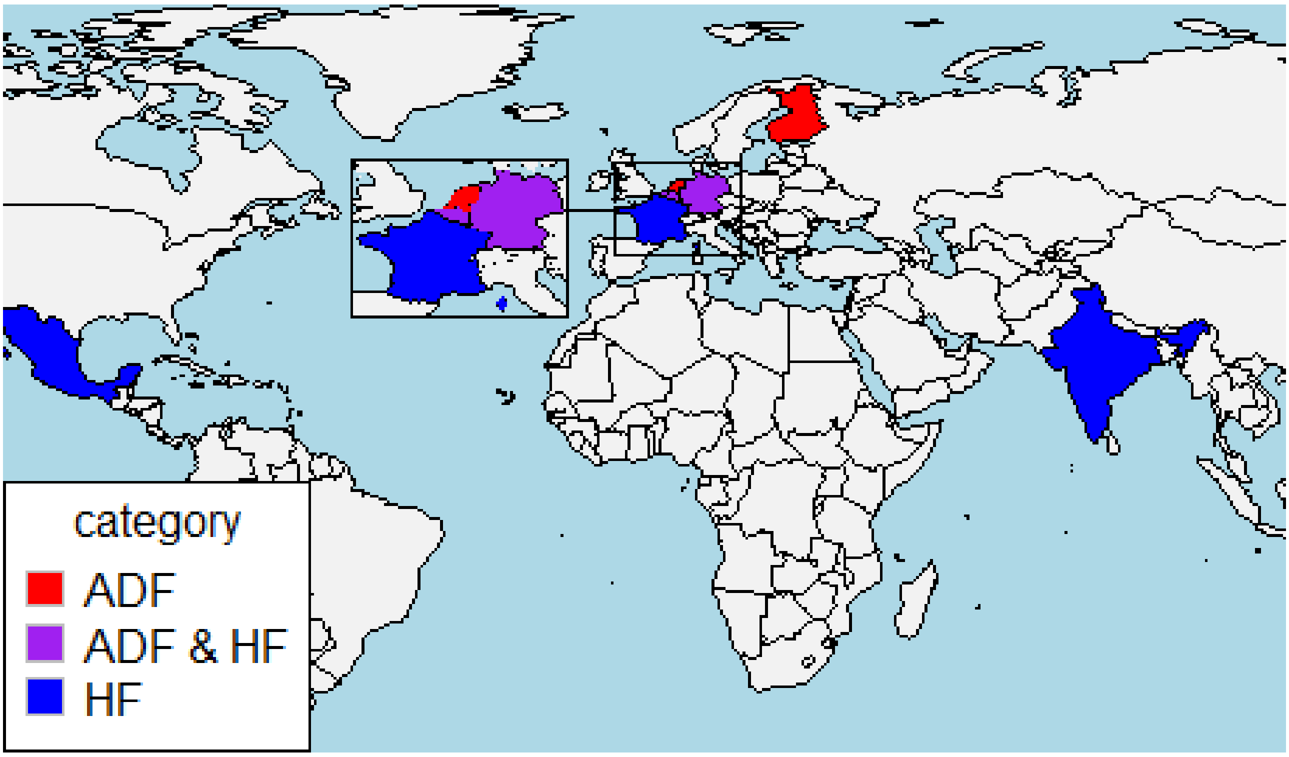


Figure 3: Largest Clubs: 1930

References

Abbott, A. and G. De Vita (2013). Testing for long-run convergence across regional house prices in the uk: a pairwise approach. *Applied Economics 45*(10), 1227–1238.

Abbott, A., G. De Vita, and L. Altinay (2012). Revisiting the convergence hypothesis for tourism markets: Evidence from turkey using the pairwise approach. *Tourism Management 33*(3), 537–544.

Apergis, N. and P. Padhi (2013). Health expenses and economic growth: convergence dynamics across the indian states. *International journal of health care finance and economics 13*(3-4), 261–277.

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

Bernard, A. and S. Durlauf (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics 71*, 161–173.

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

Bron, C. and J. Kerbosch (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM 16*(9), 575–577.

Chatterji, M. (1992). Convergence clubs and endogenous growth. *Oxford Review of economic policy 8*(4), 57–69.

Corrado, L., R. Martin, and M. Weeks (2005). Identifying and interpreting regional convergence clusters across europe. *The Economic Journal 115*(502), C133–C160.

Dufrénot, G., V. Mignon, and T. Naccache (2012). Testing catching-up between the developing countries:"growth resistance" and sometimes "growth tragedy". *Bulletin of Economic Research 64*(4), 470–508.

Durlauf, S. N. and P. A. Johnson (1995). Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics 10*(4), 365–384.

Durlauf, S. N., P. A. Johnson, and J. R. Temple (2005). Growth econometrics. *Handbook of economic growth 1*, 555–677.

Fritsche, U. and V. Kuzin (2011). Analysing convergence in europe using the non-linear single factor model. *Empirical Economics 41*(2), 343–369.

Galor, O. (1996). Convergence? inferences from theoretical models. *The Economic Journal*, 1056–1069.

Hausmann, R., L. Pritchett, and D. Rodrik (2005). Growth accelerations. *Journal of Economic Growth 10*(4), 303–329.

Hobijn, B. and P. H. Franses (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics 15*(1), 59–81.

Ikeno, H. (2014). Pairwise tests of convergence of japanese local price levels. *International Review of Economics & Finance 31*, 232–248.

Kim, Y. S. and J. J. Rous (2012). House price convergence: Evidence from us state and metropolitan area panels. *Journal of Housing Economics 21*(2), 169–186.

Konc, J. and D. Janezic (2007). An improved branch and bound algorithm for the maximum clique problem. *proteins 4*, 5.

Ozkan, H., T. Stengos, and M. E. Yazgan (2014). Persistence in convergence and club formation.

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

Pesaran, M. and A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics 10*, 461–465.

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

Quah, D. T. (1996). Empirics for economic growth and convergence. *European economic review 40*(6), 1353–1375.

Stengos, T. and M. E. Yazgan (2014). Persistence in convergence. *Macroeconomic Dynamics 18*(04), 753–782.

Yilmazkuday, H. (2013). Inflation targeting, flexible exchange rates and inflation convergence. *Applied Economics 45*(5), 593–603.