Cambridge Working Papers in Economics: 1873

ROBUST TESTS FOR CONVERGENCE CLUBS

Luisa Corrado
Thanasis Stengos
Melvyn Weeks
M. Ege Yazgan

21 December 2018

In many applications common in testing for convergence the number of cross-sectional units is large and the number of time periods are few. In these situations asymptotic tests based on an omnibus null hypothesis are characterised by a number of problems. In this paper we propose a multiple pairwise comparisons method based on an a recursive bootstrap to test for convergence with no prior information on the composition of convergence clubs. Monte Carlo simulations suggest that our bootstrap-based test performs well to correctly identify convergence clubs when compared with other similar tests that rely on asymptotic arguments. Across a potentially large number of regions, using both cross-country and regional data for the European Union we find that the size distortion which afflicts standard tests and results in a bias towards finnding less convergence, is ameliorated when we utilise our bootstrap test.

Robust Tests for Convergence Clubs[*]

Luisa Corrado[†]
University of Rome Tor Vergata and University of Cambridge

Thanasis Stengos
University of Guelph

Melvyn Weeks
Faculty of Economics and Clare College, University of Cambridge

M. Ege Yazgan
Istanbul Bilgi University

December 20, 2018

Abstract

In many applications common in testing for convergence the number of cross-sectional units is large and the number of time periods are few. In these situations asymptotic tests based on an omnibus null hypothesis are characterised by a number of problems. In this paper we propose a multiple pairwise comparisons method based on an a recursive bootstrap to test for convergence with no prior information on the composition of convergence clubs. Monte Carlo simulations suggest that our bootstrap-based test performs well to correctly identify convergence clubs when compared with other similar tests that rely on asymptotic arguments. Across a potentially large number of regions, using both cross-country and regional data for the European Union we find that the size distortion which afflicts standard tests and results in a bias towards finding less convergence, is ameliorated when we utilise our bootstrap test.

Keywords: Multivariate stationarity, bootstrap tests, regional convergence.

JEL Classifications: C51, R11, R15.

---

[*] We are grateful to Andrew Harvey, Hashem Pesaran, Jonathan Temple, Mike Wickens, Gernot Dopplehoffer, Vasco Carvalho for their comments on the work.

[†] Luisa Corrado gratefully acknowledges the Marie-Curie Intra European fellowship 039326. luisa.corrado@uniroma2.it

1 Introduction

The extent to which countries and or regions are similar across one or more dimensions is a question that has long been of interest to economists and policymakers. Within the European Union the ECB targets a single Euro Area inflation rate, and in this respect the degree to which there exists convergence in regional per capita incomes and output is of critical relevance to European regional development policies (Boldrin and Canova (2001)). Moreover, one of the core components of the European cohesion policy has been to reduce the disparities between income levels of different regions and in particular the backwardness of the least favoured regions; this objective has, in general, been manifest as the promotion of convergence between EU regions.[1] In this context it is evident that the correct detection of the extent of convergence within a regional economy is paramount given that policy usually tries to achieve regional convergence by reducing the gap between the richest and the poorest regions. In this respect a test of convergence which exhibits bias, for example being oversized in small samples, will mislead, and in this instance imply less convergence suggesting the need for more policy initiatives than may actually be required.

Economists have conceptualised the notion of similarity using formal definitions of convergence based upon growth theory. Standard neoclassical growth models (Solow (1956) and Swann (1956)) founded upon the key tenets of diminishing returns to capital and labour and perfect diffusion of technological change, dictate that countries will converge to the same level of per capita income (output) in the long run, independent of initial conditions. The New Growth theory (see, for example, Romer (1986); Lucas (1988); Grossman and Helpman (1994); Barro and Sala-i-Martin (1997)) allows for increasing returns to accumulable factors such as human capital in order to determine the (endogenous) long-run growth rate.[2] The emergence over the past decade of New Economic Geography[3] models of industrial location and agglomeration, has resulted in the identification of other forces which generate increasing returns, two notable examples being the relationship between location and transportation costs (Louveaux *et al.*, 1982) and the effect of regional externalities

---

[1] See Article 158 of the Treaty establishing the European Community.

[2] Other variants of the New Growth theories predict the emergence of multiple locally-stable steady-state equilibria instead of the unique globally-stable equilibrium of the neoclassical growth model as a result of differences in human and physical capital per worker across countries (Basu and Weil (1998)), their state of financial development (Acemoglu and Zilibotti (1997)) or other externalities caused by complementarity in innovation (Ciccone and Matsuyama (1996)).

[3] In the 'new economic geography' models the sources of increasing returns are associated with Marshallian-type external localisation economies (such as access to specialised local labour inputs, local market access and size effects, local knowledge spillovers, and the like). These models provide a rich set of possible long run regional growth patterns that depend, among other things, on the relative importance of transport costs and localisation economies (Fujita, et al. 1999; Fujita and Thisse 2002).

(Cheshire and Hay (1989)).

To the extent that the process of growth is different across regions in the sense that there are different long-run steady-states, the standard neoclassical growth model is not valid. In this context traditional approaches to test for convergence are hard to justify, difficult to interpret, and often difficult to implement. For example, a rejection of the omnibus null of convergence across a groups of regions provides increasingly less information as the number of regions increases and where prior knowledge over both the number and composition of convergence clubs is minimal. Moreover, the justification of constructing such a large intersection null hypothesis is often questionable at the outset. Faced with the emergence of larger panels, with an attendant increase in cross-sectional heterogeneity, there has been a number of significant developments in testing. For example, the use of a heterogeneous alternative hypothesis partially alleviates the problem of testing over a large group of potentially heterogenous regions (see, for example, Im et al., 2003).

In a further progression away from the testing of general omnibus hypotheses, Pesaran (2007) conducts pairwise tests for region pairs, with inference focussed on the proportion of output gaps that are stationary. One drawback of this approach is that limited inference can be made as to the significance of individual gaps, or indeed whether a group of output comparisons form a convergence club. An approach which allows for an endogenous determination of the number of clubs using a sequence of pairwise stationarity tests has been developed by Hobijn and Franses (2000). In extending this approach Corrado, Martin and Weeks (2005) developed a testing strategy that facilitates both the endogenous identification of the number and composition of regional clusters (or ‘clubs’), and the interpretation of the clubs by comparing observed clusters with a number of hypothesized regional groupings based on different theories of regional growth. However, given that the time series are relatively short, there are potential problems in basing inference on asymptotic results for stationarity tests. Reliance on large T asymptotics is likely to impart a size distortion, biasing the results towards finding less convergence than actually exists.

To circumvent this problem we propose in this paper a recursive bootstrap test for stationarity which is designed to detect multiple convergence clubs without prespecification of group membership. Monte Carlo simulations suggest that the proposed bootstrap based method performs quite well in identifying club membership when compared with the Hobijn and Franses (2000) approach that is based on asymptotic arguments. We implement our bootstrap recursive test of convergence using the original cross-country dataset used by Hobijn and Franses (2000) and the European regional data used in Corrado, Martin and Weeks (2005). We then compare the asymptotic and bootstrap generated cluster outcomes. Our results show that by resolving the size distortion which afflicts the asymptotic test we find considerably more evidence of convergence in both the applications considered.

The paper is structured as follows. Section two reviews existing tests for convergence clubs, and in section three we present the bootstrap version of the test. In section four we propose a Monte Carlo experiment to compare the properties of the asymptotic and bootstrap tests. Section five describes the data and applies the proposed tests to two real word datasets. In section six we discuss our findings and conclusions.

2 Tests for Convergence

In this section we briefly discuss a number of significant developments in tests designed to detect convergence and identify clubs which are able to address a number of questions such as whether a particular pair of countries have converged, or whether a group of regions or countries form a convergence club. We briefly discuss the different approaches to detect convergence tracking a gradual progression away from multivariate time series and panel data tests based on an omnibus null, towards sequential tests and tests that are founded upon multiple pairwise comparisons.[4] Our focus here is to identify endogenously clubs using multivariate tests for stationarity. However, given that the time series are relatively short, we show that there are potential problems in basing inference on asymptotic results for stationarity tests. To circumvent this problem we bootstrap the stationarity test and assess the effect of the size distortion on the cluster outcomes using two different applications based on country and regional level data.

The use of *multivariate time-series* to test for convergence was initiated by the seminal papers of Bernard and Durlauf (1995, 1996). Given a set $\digamma$ of N economies, a multi country definition of relative convergence asks whether the long-run forecast of all output differences with respect to a benchmark economy, (denoted with the subscript 1) tend to a country-specific constant as the forecasting horizon tends to infinity.[5] We may then write

$$\lim_{s\to\infty} E(y_{(i1),t+s} \mid I_t) = \mu_{1i} \qquad \forall i \neq 1, \tag{1}$$

---

[4] Corrado and Weeks (2011) provide a more detailed overview.

[5] A necessary condition for regions i and j to converge, either absolutely or relatively, is that the two series must be cointegrated with cointegrating vector $[1, -1]$. However, if output difference are trend stationary, this implies that the two series are co-trended as well as cointegrated. Hence a stronger condition for convergence is that output differences cannot contain unit roots or time trends (Pesaran (2007)).

where $y_{(i1),t+s} = y_{it+s} - y_{1t+s}$ and μ_{1i} is a finite constant.[6] There exist a number of problems with multivariate time series tests. First, the testing procedure is sensitive to the choice of the benchmark country. Second, in keeping with the problems of omnibus tests, in the event of rejecting the non-convergence null we have no information as to which series are $I(0)$ and $I(1)$, nor the composition of any convergence groups. Third, given the system properties of the test, a dimensionality constraint means that it can handle only a small number of economies simultaneously.

Panel unit root procedures have also been adopted to test for convergence by considering the stationary properties of output deviations with respect to a benchmark economy (Fleissig and Strauss, 2001; Evans, 1998; Carlino and Mills, 1993). First, the so called 'first-generation' panel unit-root tests,[7] maintain that errors are independent across cross-sectional units which imparts a size distortion. To overcome this problem a 'second generation' of panel unit root tests have been developed which allows for different forms of cross-sectional dependence.[8] However, as pointed out by Breitung and Pesaran (2008), panel data unit root tests poses similar problems in that as N becomes large the likelihood of rejecting the omnibus null increases with no information on the exact form of the rejection.

The problem of identifying the mix of $I(0)$ and $I(1)$ series whilst still utilising the attendant power from a panel by exploiting coefficient homogeneity under the null, has been addressed by the *sequential* test proposed by Kapetanios (2003).[9] Specifically, Kapetanios employs a sequence of unit root tests of panels of decreasing size to separate stationary and nonstationary series,[10] facilitating an endogenous identification of the number and identity of stationary series. Although a positive development there are a number of limitations. Critically the utility of this approach depends on the use of a panel framework to add power in a situation where most series are stationary but very persistent. In addition, the method only permits the classification of the N series into two groups whereas there may be many more groups. As a consequence it is not possible to address a number of questions that may be of interest: such as whether a particular pair of countries have converged, or whether a group of regions or countries form a convergence club.

When applied to output deviations, an additional problem with the Kapetanois

---

[6] We consider this as a more reasonable definition of convergence in the sense that it allows the process of convergence to stop within a neighborhood of zero mean stationarity (absolute convergence) and is consistent with the existence of increasing costs of convergence.

[7] See, for example, Maddala and Wu, 1999; Im et al., 2003; Levin, Lin and Chu, 2002.

[8] For example, Taylor and Sarno (1998) adopt a multivariate approach and estimate a system of $N-1$ ADF equations using Feasible GLS to account for contemporaneous correlations among the disturbances. Other notable example of second generation of panel unit root tests with cross-sectional dependence include Pesaran (2007) and Moon and Perron (2007).

[9] See also Flores et al. (1999) and Breuer et al. (1999).

[10] This method is referred to as the Sequential Panel Selection Method (SPSM).

test is that the testing procedure is still sensitive to the choice of the benchmark country. One approach which avoids the pitfalls of the choice of a benchmark country and, more generally, the dimensionality problem that afflicts the application of omnibus tests, is to conduct separate tests of either stationarity and/or nonstationarity. By considering a particular multi-country definition of convergence, Pesaran (2007) adopts a *pairwise* approach to test for unit-roots and stationarity properties of all $N(N-1)/2$ possible output pairs $\{y_{it} - y_{jt}\}$. The definition of convergence that is adopted is that the N countries converge if

$$\Pr(\cap_{i=1,...,N,j=i+1,...,N}|y_{it+s} - y_{jt+s}| < c|I_t) > \pi, \tag{2}$$

for all horizons, $s = 1, ..., \infty$, and c a positive constant. $\pi \geq 0$ denotes a tolerance probability which denotes the proportion that one would expect to converge by chance.

In testing the significance of the *proportion* of output gaps that indicate convergence, the dimensionality constraint that affects the application of system-wide multivariate tests of stationarity is circumvented. However, although the pairwise tests of convergence proposed by Pesaran (2007) is less restrictive than the omnibus tests proposed by Bernard and Durlauf (1995), the subsequent inference is limited in that it does not allow inference on which pairs of regions have converged, or the number and composition of convergence clubs. Below we examine various testing strategies for club convergence and in particular the *sequential testing procedure* proposed by Hobijn and Franses (2000).

2.1 Sequential Pairwise Tests

Despite the use of multivariate time series and panel data methodologies to test for convergence, there has been relatively few attempts to utilize this approach to systematically identify convergence clubs (see Durlauf *et al.* 2005). 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 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. Similarly, Tan (2009), utilizes a regression tree approach which again utilises exogenous information in the form of conditioning variables. An alternative approach to the cross-sectional notion of $\beta-$ convergence 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).

There are many other studies where the identification of convergence clubs has been achieved exogenously through testing in conjunction with a pre-classification of clubs using parametric techniques.[11] For example, Weeks and Yao (2003) adopt this approach when assessing the degree of convergence across coastal and interior provinces in China over the period 1953-1997. As Maasoumi and Wang (2008) note, the principal problem with pre-classification is that as the number of regions increases such a strategy is not robust to the existence of other convergence clubs within each sub-group. A different 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 general it is straightforward to test whether two regions form a single group. We could simply construct a single output (income) deviation and test for stationarity or a unit root. In the context of differentiating between the stationarity properties of multiple series (or output deviations), the contribution by Kapetanios (2003) has provided new techniques that utilise the power of an omnibus null in conjunction with a sequential test that allows greater inference under the alternative hypothesis. However, for a large set of regions, $\digamma$, locating partitions that are consistent with a particular configuration of convergence clubs generates difficulties given that the number of combinations is large and related, that we have little prior information.[12] Hobijn and Franses (2000) propose an empirical procedure that endogenously locates groups of similar countries (convergence clubs) utilising a sequence of stationarity tests. Cluster or club convergence in this context implies that regional per capita income differences between the members of a given cluster converge to zero (in the case of absolute convergence) or to some finite, cluster specific non-zero constant (in the case of relative convergence). Below we illustrate the method.

The Hobijn and Franses (2000) test represents a multivariate extension of the Kwiatkowski *et al.*(1992) test (hereafter KPSS test). We introduce the test by first

---

[11] See Quah (1997) for an example of non-parametric techniques to locate convergence clubs.

[12] Harvey and Bernstein (2003), utilize non-parametric panel-data methods focussing on the evolution of temporal level contrasts for pairs of economies, identifying the number and composition of clusters. Beylunioğlu et al (2018) also propose a method based on the maximum clique method from graph theory that relies on Augmented Dickey Fuller (ADF) unit root testing. However, as it will be explained below, the maximum clique method is a "top down" method that leads to a different definition of clubs and is not directly comparable to the "bottom up" Hobijn and Franses' approach.

denoting $\mathbf{y}_t = \{y_{it}\}$ as the $N \times 1$ vector of log per capita income and write $\boldsymbol{y}_t$ as

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} t + \mathbf{D} \sum_{s=1}^{t} \mathbf{v}_s + \boldsymbol{\varepsilon}_t, \tag{3}$$

where $\boldsymbol{\alpha} = \{\alpha_i\}$ is a $N \times 1$ vector of constants, $\boldsymbol{\beta} = \{\beta_i\}$ is a $N \times 1$ vector of coefficients for the deterministic trend t, and $\mathbf{v}_s = \{v_{l,s}\}$, $l = 1, ..., m$ represents a $m \times 1$ vector of first differences of the m stochastic trends in $\mathbf{y}_t$, $m \in (0, ..., N)$. $\mathbf{D} = \{D_{i,l}\}$ denotes a $N \times m$ matrix with element $D_{i,l}$ denoting the parameter for the l^{th} stochastic trend. $\boldsymbol{\varepsilon}_t = \{\varepsilon_{i,t}\}$ is a $N \times 1$ vector of stochastic components.

In considering the difference in log per capita income for regions i and j we write

$$y_{(ij),t} = \alpha_{(ij)} + \beta_{(ij)} t + \sum_{l=1}^{m} D_{(ij),l} \left(\sum_{s=1}^{t} v_{l,s} \right) + \varepsilon_{(ij),t}, \tag{4}$$

where (4) admits two different convergence concepts: absolute and relative. The restrictions implied by the null of relative convergence are $\beta_{(ij)} = 0 \;\; \forall i \neq j \in \digamma$ and $D_{(ij),l} = D_{il} - D_{jl} = 0 \;\; \forall l = 1, ..., m$, with the latter restriction indicating that the stochastic trends in log per capita income are cointegrated with cointegrating vector $[1 - 1]$. The additional parameter restrictions for the null hypothesis of absolute convergence are that $\alpha_{(ij)} = 0 \;\; \forall i \neq j \in \digamma$. Both asymptotic absolute and relative convergence imply that the cross sectional variance of log per capita income converges to a finite level.

Denoting the partial sum process $S_t = \sum\limits_{s=1}^{t} y_{(ij),s}$, the test statistic for zero mean stationarity is given by[13]

$$\widehat{\tau}_0 = T^{-2} \sum_{t=1}^{T} S_t^{'} [\widehat{\sigma}^2]^{-1} S_t. \tag{6}$$

Denoting $h_t = y_{(ij),t} - \frac{1}{T} \sum\limits_{t=1}^{T} y_{(ij),t}$ and the partial sum process as $\bar{S}_t = \sum\limits_{s=1}^{t} h_s$ the test

---

[13] For the stationary null, Hobijn and Franses (2000) utilise the Kwiatkowski *et al.* (1992) test. The KPSS test is operationalised by regressing the pairwise difference in per capita income $y_{(ij),t}$ against an intercept and a time trend giving residuals

$$\widehat{\varepsilon}_{(ij),t} = y_{(ij),t} - \widehat{\alpha}_{(ij)} - \widehat{\beta}_{(ij)} t. \tag{5}$$

and

$$\widehat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}^2_{(ij),t} + 2 \frac{1}{T} \sum_{k=1}^{L} \omega(k, L) \sum_{t=k+1}^{T} \widehat{\varepsilon}_{(ij),t} \widehat{\varepsilon}_{(ij),t-k},$$

represents the consistent Newey-West estimator of the long-run variance. $\omega(k, L) = 1 - k/(1 + L)$, $k = 1, ..., L$ is the Bartlett kernel, where L denotes the bandwidth.

statistic for level stationarity is given by

$$\widehat{\tau}_{\mu} = T^{-2}\sum_{t=1}^{T}\bar{S}_t^{'}[\widehat{\sigma}^2]^{-1}\bar{S}_t. \tag{7}$$

Examining (4), we note that in the case of two regions, and focussing on a test of relative convergence with restrictions $D_{(ij),l} = 0 \ \forall l = 1, ..., m$ and $\beta_{(ij)} = 0$, it is obviously straightforward to test whether two regions form part of a single group. However, for a large number of regions locating the partitions over $\mathcal{F}$ that are consistent with a particular configuration of convergence clubs is infeasible both because the number of combinations is large and related, that we have little prior information on the form of $\mathbf{D}$ and the likely combination of zeros restrictions over the differences $\beta_{(ij)}$ and $\alpha_{(ij)}$. The alternative testing strategy proposed by Hobijn and Franses (2000) forms groups from the bottom up using a clustering methodology to determine, endogenously, the most likely combination of restrictions, and as a consequence, the most likely set of convergence clubs. The cluster algorithm is based on the hierarchical farthest neighbour method due to Murtagh (1985). We illustrate the sequential test using the set of regions $\mathcal{F} = \{1, 2, 3, 4\}$.

(i) We first initialise singleton clusters $K(i)$ for each region $i = 1, ..., 4$. The null hypothesis of level stationarity is tested for all $N(N-1)/2 = 6$ region pairs. We collect p-values in the vector $\widehat{\mathbf{p}}_{s=1} = \{p_{(ij)}\}$, where $p_{(ij)} = \Pr(\widehat{\tau}_{(ij),\mu} < c_{(ij)}|I_t)$, $\widehat{\tau}_{(ij),\mu}$ denotes the test statistic and $c_{(ij)}$ the critical value. $s = 1$ denotes the first iteration.

Clusters are formed on the basis of the max p-value in $\widehat{\mathbf{p}}_{s=1}$, indicating the pair of regions which are most likely to converge. If, for example, $p_{(1,2)} = \{\max_{i,j\in\mathcal{F}}\{\widehat{\mathbf{p}}_{s=1}\} > p_{\mathtt{min}}\}$ then regions 1,2 are the first pair of regions to form a club.[14] We denote the first cluster as $K(1^{'}) = \{1, 2\}$ and discard the singleton cluster 2, which is now part of the two-region cluster $K(1^{'})$.

(ii) In the second iteration ($s = 2$) we define the set of regions as $\mathcal{F}^{'} = (1^{'}, 3, 4)$. We form pairwise output differences between the $N - 2$ remaining singleton clusters and the two-region cluster $K(1')$. Once again we collect the p-values in the vector $\widehat{\mathbf{p}}_{s=2}$. Letting $p_{(r,v)} = \{\max_{i,j\in\mathcal{F}}\{\widehat{\mathbf{p}}_{s=2}\} > p_{\mathtt{min}}\}$, then if, for example, $p_{(r,v)} = p_{(1'3)}$, the singleton cluster $K(3)$ joins cluster $K(1')$ forming a three-region cluster $K(1^{''}) = \{1, 2, 3\}$.

---

[14] The choice of $p_{\min}$ has a direct effect on the cluster size. Since the stationarity test is known to be oversized in small samples, this bias will generate inference towards finding less convergence.

(iv) In this example we find a three-region cluster $K(1'')$ and a singleton cluster $K(4)$, so the procedure stops.

The principle difference between this sequential testing strategy and the SPSM approach of Kapetanios is that the SPSM test is designed to endogenously classify stationary and nonstationary series. This is achieved by sequentially reducing the size of the omnibus 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. In contrast the Hobijn and Franses method seeks to endogenously allocate N series to $J \leq N$ convergence clubs. This is achieved by only classifying regions that provide, at each recursion and conditional on exceeding $p_{\mathtt{min}}$, the most evidence for convergence.

Although the sequential multivariate stationarity test is consistent in that for large T the tests will reveal the true underlying convergence clubs, the principle shortcoming is that the test statistic is known to be oversized in small samples (Caner and Kilian, 2001). When testing for convergence using yearly data T is likely to be small, and as a result inference is likely to be biased in the direction of finding less convergence. Similar size distortions also emerge when the series are stationary but highly persistent: in this case the partial sum of residuals which are used to derive the KPSS test resemble those under the alternative in the limit. Below we outline a bootstrap approach which circumvents the pitfalls of inference based upon asymptotic arguments since it is able to generate independent bootstrap resamples using a parametric model which is conditional on the sample size and the dependence structure of the dataset. In section 5 we utilise this test to investigate the extent of convergence in two different applications: (i) the cross-country dataset originally adopted by Hobijn and Franses (2000); (ii) the European regional dataset used by Corrado, Martin and Weeks (2005).

3 A Bootstrap Test

To derive the parametric model with which to create independent bootstrap samples under the stationarity null, following Kuo and Tsong (2005) and Leybourne and McCabe (1994), we exploit the equivalence in second order moments between an unobserved component model and a parametric ARIMA model (Harvey (1989)) for the differenced data. In demonstrating this equivalence we note that (4) may be rewritten in structural form as a function of a deterministic component $(\alpha_{(ij)} + \beta_{(ij)}t)$, a random walk ($r_t$) and a stationary error ($\varepsilon_{(ij),t}$):

$$\begin{aligned} y_{(ij),t} &= \alpha_{(ij)} + \beta_{(ij)} t + \sum_{l=1}^{m} D_{(ij),l} r_{l,t} + \varepsilon_{(ij),t} & (8) \\ r_{l,t} &= r_{l,t-1} + v_{l,t}, & (9) \end{aligned}$$

where $r_{l,t} = \sum_{s=1}^{t} v_{l,s}$ represents the l-th stochastic trend for regions i and j with r_0, the fixed initial value, set to zero. We also assume that $\varepsilon_{(ij),t}$ is a stationary error process $\varepsilon_{(ij),t} = \sum_{s=0}^{\infty} \psi_{(ij),s} u_{(ij),t-s} = \Psi(L) u_{(ij),t}$ where $\psi_{(ij),0} = 1$, $u_{(ij),t} \sim i.i.d(0, \sigma^2_{u_{(ij)}})$ and $\Psi(L) = 1 + \sum_{s=1}^{\infty} \psi_{(ij),s} L^s$.[15] Under these assumptions $\left\{\varepsilon_{(ij),t}\right\}$ has an infinite order autoregressive representation

$$\varepsilon_{(ij),t} = \sum_{s=1}^{\infty} \lambda_{(ij),s} \varepsilon_{(ij),t-s} + u_{(ij),t}, \qquad (10)$$

where $\Lambda(L) = \Psi(L)^{-1} = 1 + \sum_{s=1}^{\infty} \lambda_{(ij),s} L^s$. Given (8), since $\left\{\varepsilon_{(ij),t}\right\}$ is a stationary process, the necessary condition for convergence of regions i, j is that the variance of the random walk error (σ^2_v) is zero. Focussing on a test for relative convergence, below we describe the nature of the recursive multivariate stationarity test using critical values generated from the empirical distribution of the test statistic constructed using bootstrap sampling.

In generating a bootstrap test for relative convergence we focus on relative convergence where $\beta_{(ij)} = 0$, which rules out the presence of a deterministic trend.[16] The idea is to estimate the null finite sample distribution of the KPSS test statistics by exploiting the equivalence between the unobservable component model and the parametric ARIMA model. Harvey (1989) demonstrates that the components from the structural model (8) can be combined to give a reduced form ARIMA(0,1,1) model. In particular, assuming independence of $\varepsilon_{(ij),t}$ and v_t, (8) becomes a local component model which, after time differencing, can be expressed as the MA model $\Delta y_{(ij),t} = (1 - \theta L)\eta_{(ij),t}$ where $\eta_{(ij),t} \sim i.i.d(0, \sigma^2_{\eta_{(ij)}})$ and $\sigma^2_{\eta_{(ij)}} = \sigma^2_{\varepsilon_{(ij)}}/\theta$.

The reduced form parameter θ is derived by equating the autocovariances of first differences at lag one in the structural and reduced forms. This gives the following relationship between the parameters of the components model (8) and the

---

[15] $\varepsilon_{(ij),t}$ is assumed to be invertible $\sum_{s=0}^{\infty} s \left|\psi_{(ij),s}\right| < \infty$.

[16] For the test of absolute convergence the restrictions are $\beta_{(ij)} = \alpha_{(ij)} = 0$.

ARIMA(0,1,1) model:

$$\theta = \frac{1}{2}\left\{\frac{\sigma_v^2}{\sigma_{\varepsilon_{(ij)}}^2} + 2 - \left(\frac{\sigma_v^2}{\sigma_{\varepsilon_{(ij)}}^2} + 4\frac{\sigma_v^2}{\sigma_{\varepsilon_{(ij)}}^2}\right)^{1/2}\right\}. \tag{11}$$

where $q = \frac{\sigma_v^2}{\sigma_{\varepsilon_{(ij)}}^2}$ is the signal to noise ratio. Under the stationarity null, namely that regions i, j are converging, the variance of the random walk component (σ_v^2) is zero, which in turn implies that $\theta = 1$ in the ARIMA representation. Therefore by imposing a moving average unit root in the ARIMA representation one can use the parametric model for sampling instead of the "unobservable" component model.

Our bootstrap sampling scheme is based on the following procedure. First, for each region pair i, j and contemporaneous difference $y_{(ij),t} = y_{i,t} - y_{j,t}$, we fit an ARMA$(p, 1)$ model to the differenced series $\Delta y_{(ij),t} = y_{(ij),t} - y_{(ij),t-1}$, namely

$$\Delta y_{(ij),t} = \sum_{k=1}^{p} \phi_{(ij),k} \Delta y_{(ij),t-k} + \eta_{(ij),t} - \theta \eta_{(ij),t-1}, \tag{12}$$

The MA component in (12) follows from the reparametrisation of the structural component model to reproduce the stationarity properties of the data in the ARMA representation. The AR(p) component[17] represents an approximation to the assumed infinite-order moving average errors to capture the dependence structure in the data. By imposing a moving average unit root in the sampling procedure we can then construct the bootstrap distribution of the test statistic for level stationarity defined in (7).

The accuracy of the bootstrap test relative to the asymptotic approximations hinges on the bootstrap sample being drawn independently. Given the presence of a known dependence structure, in this case a stationary ARMA$(p, 1)$ model, we utilise the Recursive Bootstrap.[18] To achieve independent re-sampling from (12) we estimate $\hat{\phi}_{(ij),k}$ and $\hat{\eta}_{(ij),t}$, and we draw a bootstrap sample $\{\bar{\eta}^r_{(ij),t}\}_{t=1}^T$ from the distribution of centered[19] residuals $\{\bar{\eta}_{(ij),t}\}_{t=1}^T$, where $\bar{\eta}_{(ij),t} = \widehat{\eta}_{(ij),t} - \frac{1}{T-1}\sum_{t=2}^T \widehat{\eta}_{(ij),t}$.

Given the bootstrapped residuals, $\{\bar{\eta}^r_{(ij),t}\}_{t=1}^T$, the r^{th} bootstrap sample for the

---

[17] p denotes optimal lag length, chosen using the AIC criterion.

[18] See Horowitz (2001) on the merits of the recursive bootstrap for linear models, and Maddala and Li (1997) and Efron and Tibshirani (1986) for specific examples.

[19] Centering the residuals reduces the downward bias of autoregression coefficients in small samples (Horowitz, 2001).

data $\{\Delta y^r_{(ij),t}\}_{t=1}^T$ is generated based on the recursive relation[20]

$$\Delta y^r_{(ij),t} = \sum_{k=1}^{p} \widehat{\phi}_{(ij),k} \Delta y^r_{(ij),t-k} + \bar{\eta}^r_{(ij),t} - \bar{\eta}^r_{(ij),t-1}. \tag{13}$$

We then recover the level of the series (where the level denotes the contemporaneous regional difference) directly from (13)

$$y^r_{(ij),t} = y^r_{(ij),t-1} + \sum_{k=1}^{p} \widehat{\phi}_{(ij),k} \Delta y^r_{(ij),t-k} + \eta^r_{(ij),t} - \eta^r_{(ij),t-1}. \tag{14}$$

Defining $h^r_t = y^r_{(ij),t} - \frac{1}{T} \sum\limits_{t=1}^{T} y^r_{(ij),t}$, then for r^{th} bootstrap sample, and the i, j region pair, a test statistic for relative convergence, $\widehat{\tau}^r_{(ij),\mu}$, is given by

$$\widehat{\tau}^r_{(ij),\mu} = T^{-2} \sum_{t=1}^{T} \bar{S}^{r,\prime}_t [\widehat{\sigma}^{r,2}]^{-1} \bar{S}^r_t, \tag{15}$$

where $\bar{S}^r_t = \sum\limits_{s=1}^{t} h^r_s$. For each region pair we draw R bootstrap samples and construct the empirical distribution of the test statistic under the null, which we denote $\tau^B_{(ij),\mu}$. Bootstrap critical values $C^B_{(ij),\mu}$ can then be recovered at the required significance levels and we can implement the algorithm described in section 3 utilising a vector of bootstrapped empirical p-values, $\hat{\mathbf{p}}^B$.

4 A Monte Carlo Study

In this section we compare the performance of the bootstrap version of the KPSS test, CW henceforth, with the original HF test based on the asymptotic version of the multivariate KPSS testing procedure. To the best of our knowledge there is no other comparable Monte Carlo study in the literature that evaluates clustering methods in the same context as we do here.

HF is a method that relies on a "bottom up" algorithm that clusters groups one by one. To determine whether a set of countries is convergent, HF applies a multivariate stationarity test to panels comprised of consecutive pairwise difference series set elements and confirms convergence if the null hypothesis of stationarity

---

[20] Initial values, $\Delta y^r_{(ij),t-1} = y^r_{(ij),t-1} - y^r_{(ij),t-2} = ... = \Delta y^r_{(ij),t-p} = y^r_{(ij),t-p} - y^r_{(ij),t-p-1}$ are set to zero.

of the panel is not rejected using the KPSS test. For example, if we want to test the convergence of countries 1,2,3 and 4, a panel consisting of y^{12}, y^{13}, y^{23}, y^{14}, y^{24} and y^{34} is subjected to the KPSS test, where for example y^{12} ,y^{23}, and y^{34} denote the difference in log per capita between countries 1 and 2, 2 and 3 and 3 and 4 respectively. If stationarity cannot be rejected the panel is then augmented with series other than 1, 2, 3 and 4, each added separately. If then for each of these additional panels the stationarity null is rejected, then these four countries are said to be convergent.

4.1 Monte Carlo Structure

In this subsection, we will discuss the data generating processes that is used in our Monte Carlo study. We generated a number of datasets to conduct the evaluation of the clustering methods CW and HF that we compare. We will examine the performance of these methods to determine success rates in detecting club membership for various parameter configurations including the number of countries, club size, time span and number of clubs. The analysis is carried out for two separate cases. In the first case we analyze single club data, while in the second case we include multiple clubs.

Below we present the data generating processes and evaluation procedures employed in this study. A similar design was used by Beylunioğlu et al (2018) in assessing the properties of the maximum clique method, an alternative clustering mechanism that relies on Augmented Dickey Fuller (ADF) unit root testing. However, the maximum clique method is a "top down" method that leads to a different definition of clubs and is not considered in the present comparison.[21]

Data Generating Processes

The simulation assumes that the log GDP series for region i is given by

$$y_{it} = \alpha_i + d_i r_t + \epsilon_{it}, \tag{16}$$

where $\epsilon_{it} \sim I(0)$ is the error term and r_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 the country specific constants, d_i,

---

[21] 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 reason that it is based on the notion of σ 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 CW method developed in this study as both of the latter deal with convergence of the mean (function) of the series.

which measure the impact of common factor are equal. In other words, for the pair i and j, if $i = d_j$, r_t is canceled out and $y_{it} - y_{jt}$ becomes $\alpha_i - \alpha_j + \epsilon_{it} - \epsilon_{jt}$. In this case, since the error terms are assumed to be stationary, we have $\alpha_i - \alpha_j + \epsilon_{it} - \epsilon_{jt} \sim I(0)$ and the pair i and j would be convergent by definition. Likewise, for any subset of countries having equal d_i, all pairwise difference series in that subset would be stationary and hence these countries would constitute a convergence club. Finally, the constants, α_i are country specific and are generated once for all data sets.

The non-stationarity of r_t is modeled using the ARIMA process

$$r_t = r_{t-1} + v_t, \quad v_t = \rho_v v_{t-1} + e_t, \quad e_t \sim iid\ N(0, 1 - \rho_v^2), \tag{17}$$

where we allow $\rho_v = \{0.2, 0.6\}$ as separate cases. In addition we also allow the error term of the log GDP series in equation (16) to have serial dependence, following the specification

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + v_{it}, \quad v_{it} \sim \ iid\, N(0, \sigma_{v_i}^2(1 - \rho_i^2)), \tag{18}$$

where we assume that the error terms v_{it} are i.i.d. distributed Normal random variables. The autoregressive coefficient ρ_i and $\sigma_{v_i}^2$ are country specific and invariant among the single and multiple clubs datasets. 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 dataset containing a single club the coefficients of the m convergent countries are assumed to be $d_i = d_j = 1$. For the remaining $(N - m)$ countries, d_i is generated randomly as $d_i \sim iid\ \mathcal{X}_m^2$. Similarly, we also generate country specific constants as $\alpha_i \sim iid\ \mathcal{X}_m^2$.

For multiple clubs, in order to assess successful detection in club membership we want to make sure that there are some non-convergent countries present in the data that do not belong to any club. In that case, the value m of club size, when the number of clubs (k) and the number of countries (N) are given, is chosen in such a way as to allow for at least a pair of non-convergent countries to be present in order to evaluate successful converging behaviour. For a given k and N, the clubs sizes m's are randomly drawn from a Poisson distribution with a rate of $\lambda = N/k$. For each N, random draws are repeated k times.[22]

The simulations are repeated 2000 times using different combinations of $T =$

---

[22] Obviously we did not allow the sum of m to exceed N, if this happens we redraw the last club size.

$\{50, 100\}$ time intervals, $N = \{10, 20, 30\}$ count of countries, $m = \{3, 5, 7, 10\}$ number of club members for the single club case. In the multiple club case we considered $T = \{50, 100\}$, $N = 10$, $m = \{3, 5\}$ and $k = \{2, 3\}$ number of clubs.[23]

4.1.1 Testing and Evaluating Procedures

To evaluate convergence we utilise evaluation tools from the literature on forecasting. The first one is the Kupiers Score (KS), while the second one is the Pesaran and Timmermann (1992) (PT) test statistic commonly used in the forecasting times series literature for the evaluation of sign forecasts. It is worth noting that sign forecasts are used for predicting whether an underlying series would increase relative to a benchmark such as, for example, a zero return threshold. This test is cast in terms of a binary process where success is the increase relative to the chosen benchmark. In our case we take a "success" as the correct detection of a country's membership in a club. In the context of forecasting, this is equivalent to success in forecasting the sign of a time series.

Since granting membership into a club or denying it can occur randomly,[24] KS takes the correct forecasts and false alarms into account separately. KS is defined as $H - F$ where

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

I (O) are binary indicators indicating whether the country under investigation is a member (not a member) of a given club. In considering the pairs of letters, the first letter indicates whether the country is found to be a member in the Monte Carlo experiment, while the second letter denotes its actual membership state (i.e. whether the country is actually in the club or not). II then indicates that a country as a member of the club is correctly identified; OO denotes that a country is correctly identified as not a member of the club. Furthermore, IO indicates that a country is detected to be a member of the club, while actually it is not (false detection). OI refers to the reverse case where the country is misclassified 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

---

[23] The computational burden for larger values of m, k and N proved to be be quite high at this point.

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

significance of KS, we will employ the following PT statistic

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

$\widehat{P}$ 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 independent (where success is a random event of probability 0.5). $\widehat{V}(\widehat{P})$ 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 either the KS or the PT statistic given that the success/failure classification is no longer binary - as in the case of the single club case. In the multiple club case there are more than two distinct cases for the actual membership state: the country can be either a member of the correct club, belong to the "wrong" club, or not be a member of any club.

To confront this problem, in the case of multiple clubs we utilise a much stricter criterion by counting the successful cases in our simulations in which *all* countries are detected correctly. We do not evaluate success as a binary outcome, country by country as in the case of a single club in each replication, but we only count as success having all countries satisfying the convergence condition. This is a much stricter criterion given that success depends on the overall results in each replication in which *all* countries are detected correctly.

4.2 Simulation Results

Below we discuss the findings of the simulations based on the data generating processes of club formation. The comparison involves the bootstrap version of the KPSS test (henceforth CW) proposed in this paper, and the original HF test based on the asymptotic version of the multivariate KPSS test.

4.2.1 Single Club Results

The results are presented in Table 1 for 0.05 and 0.10 significance levels.[25] The total number of countries N are set at ($N = 10, 20, 30$); there are two choices of time span ($T = 50, 100$) that mimic the real data time span availability; and two choices of the persistence parameter ($\rho_v = \{0.2, 0.6\}$). It is expected that as the number of

---

[25] We also have the results for the 0.01 significance level but to conserve space we do not report them. They are available from the authors on request.

countries N and club size m increase, the likelihood of an incorrect classification will also increase, but the opposite will be the case for an increase of the time span T for given N and m.

As seen in Table 1, CW outperforms HF in all categories. For example, with $m = 3$, $N = 10, T = 50$ and $\rho = 0.2$ (configuration $\mathcal{A}$), and significance levels 0.05 and 0.10, the KS results (the "correct hit" ratio net of "false alarms") for CW are, respectively, 0.87 and 0.89. The comparable numbers for HF are 0.63 and 0.66. Similarly, for the cases with $m = 10$, $N = 30, T = 100$ and $\rho = 0.6$ (configuration $\mathcal{B}$), CW with 0.77 and 0.76 outperforms the HF method - 0.60 and 0.60. The results are in line with our prior expectations that larger N and m values would result in lower success rates. However, in all cases the CW test does better.

The PT statistics[26] for configuration $\mathcal{A}$ yield values 1.91 and 2.02 for HF and 2.51 and 2.49 for CW; for configuration $\mathcal{B}$ the HF values are 2.86 and 2.97; with 3.37 and 3.48 for CW.

Note that the rejections of the null hypothesis of random success outcomes are higher with the PT test for CW in all cases. The results clearly demonstrate that the CW, bootstrap KPSS test offers a significant improvement over the HF procedure.

4.2.2 Multiple Club Results

The results for the multiple club case are presented in Table 2. The multiple clubs cases involve classifications with $k = 2$ and 3 and $N = 10$ for $T = \{50, 100\}$ and $\rho_v = \{0.2, 0.6\}$. The club sizes associated with each club are listed in the second column of Table 2 for each k. For example, the entry $4, 4$ for club size m refers to two clubs of equal size 4, for $k = 2$. In the case of $k = 3$, m enters as $3, 3, 2$, that is two clubs of size 3 and one club of size 2.

The CW test outperforms HF in the majority of cases. For example, with $N = 10$, $k = 2$, $T = 100$ and $\rho_v = 0.6$, CW detects 54%, 45% and 37% correct classifications at the 0.01, 0.05 and 0.10 significance levels; HF does that with frequency 44%, 38% and 34.80% respectively. For the case when $k = 2$ and $T = 50$, HF does slightly better than CW, but when $k = 3$ the performance of HF deteriorates rapidly. In that case when $N = 10$, $k = 3$, $T = 100$ and $\rho_v = 0.2$, CW detects 27%, 25% and 22% correct classifications, while the comparable results for HF, respectively, are 10.40%, 8.20% and 5.60%. Since we have adopted a much stricter criterion where success is defined as *all* countries detected correctly, we do expect lower rates of correct detection than was the case for single clubs. In all cases, we see an improvement for CW when the number of clubs increases even when T is relatively small, but not for HF.

---

[26] The PT statistic follows an asymptotic standard normal variate.

Overall, the multiple club results suggest that in terms of accuracy the CW does better in detecting the presence of clubs or clusters of countries. This gives us confidence that applying the above method to real data can provide us with useful insights about how countries over time collect themselves into different club formations of similar characteristics as far as economic activity is concerned.

5 Applications

As shown in the Monte Carlo study a problem with the asymptotic test is that it does not permit reliable inference with only 30 years of data. Using an asymptotic test of the null of stationarity (convergence) tends to distort club membership detection due to size distortions which are ameliorated when we implement the bootstrap. That is, the test is oversized resulting in a tendency to reject the null hypothesis of convergence. In this section we assess the extent to which a size distortion affects our inference on the degree of convergence using two real-world datasets. We first compare the results of the asymptotic and bootstrap tests using the cross-country dataset originally adopted by Hobijn and Franses (2000). We then utilise data gathered at a finer geographical and sectoral scale by making the same comparison based upon the European regional dataset used by Corrado, Martin and Weeks (2005). We expect to resolve the size distortion which afflicts the asymptotic test and to find more evidence of convergence than what originally acknowledged by Hobijn and Franses (2000).

5.1 Cross-Country Convergence

In this section we compare the asymptotic and the bootstrap results using the Hobijn and Franses (2000) dataset for the period 1960-1989 which comprises 112 countries from the Penn World Table listed in Table 3. Focussing upon log per capita GDP, the results based upon the asymptotic test have a striking feature, namely a very large number of convergence clubs. In particular, Hobijn and Franses (2000) find 63 asymptotically perfect convergence clubs and 42 asymptotically relative[27] convergence clubs.[28] As Table 4 shows, in the case of perfect convergence the lack of convergence is manifest in 29 singletons and 22 two-country clusters. A similar result can be observed in the case of relative convergence where Hobijn and Franses (2000) find a large number of two and three-country clusters. The lack of convergence is also

---

[27] Note that perfect convergence implies convergence to identical log real GDP per capita levels. Relative convergence implies convergence to constant relative real GDP per capita levels.

[28] These are the results using $p_{\min} = 0.01$ and $L = 2$ as presented in Tables BII and BIII of the Hobijn and Franses (2000) paper.

evident in the fact there are no clusters of size six or more for asymptotic perfect convergence and only one club of size six for relative convergence.

We therefore implement the bootstrap version of the test on the same dataset and find a significant increases in the extent of convergence. Specifically, in the perfect convergence case and relative to the asymptotic results, we observe a 57% reduction in the number of convergence clubs (from 63 to 27); for relative convergence the reduction is 38% (from 42 to 26). In other words, there is evidence towards finding more convergence.

Looking at the change in the distribution of cluster sizes for perfect convergence, we observe a dramatic reduction (by 96%) in the number of singletons (from 29 to 1) and by 81% in the number of two-country clusters (from 22 to 4). Commensurate with this finding, we note that countries are now clustering at a larger scale with two clubs having up to seven countries and with a substantial increase in the number of clusters containing five and six countries. A similar increase in the degree of convergence can be observed in the case of relative convergence.

Tables 5 and 6 report the asymptotic and bootstrap cluster composition for relative convergence. A number of noteworthy observations can be made. We confirm the findings of Hobijn and Franses that convergence is more widespread among low income economies, and in particular Sub-Saharan Africa. Similarly we find that in general low income countries do not converge to high income. The two exceptions to this found in the asymptotic results, namely Kenya and Ecuador forming clubs alongside Australia and Denmark and Canada, are not found in the bootstrap results. In contrast to Hobijn and Franses we do find a significantly higher degree of convergence, both amongst low income and high income countries. The results based on the asymptotic test indicate very little convergence for the richer economies with all groups of size two. The bootstrap test locates a greater degree of convergence, for example, cluster 11 (Germany, Denmark, France, Luxemburg and New Zealand) and cluster 17 (Belgium, Great Britain, Netherlands and Norway).

In the next section we apply the asymptotic and bootstrap version of the test to the European regional dataset originally used by Corrado, Martin and Weeks (2005). A critical difference with respect to the analysis undertaken at the aggregate country level, is that we allow for the possibility that convergence is more prevalent at a sector-specific level, and in addition consider a smaller geographical unit. Much of the theory of convergence highlights the potential role of technology spillovers as one of the possible drivers of convergence. As a result, in what follows we move away from an aggregate analysis to considering how convergence differs across agricultural, manufacturing and service sectors.

5.2 European Regional Convergence

In the following sections we examine the extent of regional convergence within the EU. Regional convergence – or what the European Commission calls 'regional cohesion' – is a primary policy objective, and is seen as vital to the success of key policy objectives, such as the single market, monetary union, EU competitiveness, and enlargement (European Commission, 2004). As a result, the theory of and evidence on long-run trends in regional per capita incomes and output are of critical relevance to the EU regional convergence and regional policy debate (Boldrin and Canova, 2001). Indeed, according to Fujita *et al.* (1999), the implications of increasing economic integration for the EU regions has been one of the factors behind the development of the 'new economic geography' models of regional growth. To date, however, very few of these models have been tested empirically on EU evidence.

In response to the policy and research questions outlined above our empirical analysis will be framed around the identification of regional convergence clubs in the EU. To identify regional convergence clusters we use the method introduced by Hobijn and Franses (2000) which allows for the endogenous identification of the number and membership of regional convergence clusters (or 'clubs') and compare the results of the bootstrap and asymptotic versions of the test to assess the differences in terms of number, size and composition of the resultant clusters.

5.2.1 Data

The so-called Nomenclature of Statistical Territorial Units (NUTS) subdivides the economic territory of the 15 countries of the European Union using three regional and two local levels. The three regional levels are: NUTS3, consisting of 1031 regions; NUTS2, consisting of 206 regions; and NUTS1 consisting of 77 regions. NUTS0 represents the delineation at the national level and comprises France, Italy, Spain, UK, Ireland, Austria, Netherlands, Belgium, Luxemburg, Sweden, Norway, Portugal, Greece, Finland, Denmark and West Germany. We are aware of the problems that surround the choice of which spatial units to use.[29] For example, many of the regional units used by EUROSTAT have net inflows of commuters and in addition, these regions also tend to be those with the highest per capita income. Boldrin and Canova (2001) criticize the European Commission for utilizing inappropriate regional units. Whereas NUTS1, NUTS2 and NUTS3 regions are neither uniformly large or sufficiently heterogeneous such that a finding of income divergence across regions cannot unequivocally be taken as evidence for the existence of an endogenous

---

[29] Chesire and Magrini (2000) provide a useful discussion of these issues, focussing on the importance of centering the analysis on regions that are self-contained in labour market terms.

cumulative growth processes. In fact, the smaller the geographical scale, the more incomplete and fragmented is the statistical information we can get. Although we do not wish to detract from the importance of these matters, in this study our primary focus is a comparison of two different tests for regional convergence for which the unit of analysis is the same. In conducting our analysis we choose to focus on NUTS1 regions, achieving a compromise between the availability of reliable data at a regional level which is sufficiently homogeneous, and the need to move beyond national borders. The complete list of NUTS1 regions[30] used in this study is given in Table 7.

We use regional data on Gross Value Added[31] per worker for the period 1975 to 1999 for the agriculture, manufacturing and services sectors. Although data are available for more recent years, we focus on this particular time frame to facilitate a comparison with the results of Corrado, Martin and Weeks (2005). The service sector has been further sub-divided into market and non-market services: market services comprise distribution, retail, banking, and consultancy; non-market services comprise education, health and social work, defence and other government services.

5.2.2 Results

In this section we present the main results of our analysis. Given the large number of EU regions in Figures 1 and 2 we first present the results for the asymptotic and bootstrap test of convergence in mapped rather than tabular form. Table 8 summarises this information in terms of the number and size of the convergence clubs and group characteristics, such as average per-capita income.

5.2.3 Graphing Convergence Clusters

In Figures 1 and 2 clusters which contain the largest number of member regions are indicated with a darker shade on each map. Regions which belong to two-region clusters or do not cluster with any other region have no shading. In the key to the maps, the first number indicates the cluster size and the second letter denotes the cluster identifier. In Figure 1 maps a) and b) (c) and d)) present the asymptotic and

---

[30] For Portugal, Luxemburg and Ireland, data are only available at the NUTS0 level. For Norway we have no data at the NUTS1 level. Time series data for the sample period considered are not available for East Germany, which is therefore excluded from the analysis.

[31] GVA has the comparative advantage with respect to GDP per capita of being the direct outcome of various factors that determine regional competitiveness. Regional data on GVA per-capita at the NUTS1 level for agriculture, manufacturing, market and non-market services, have been kindly supplied by Cambridge Econometrics, and are taken from their European Regional Database. All series have been converted to constant 1985 prices (ECU) using the purchasing power parity exchange rate.

bootstrap generated outcomes for agriculture (manufacturing). The relative pattern of convergence corroborates with our prior expectations, namely that the bootstrap test is obviously rejecting the stationary null with a lower frequency and thereby locating more evidence for convergence. In Figure 2 we find a similar pattern for market and non-market services.

In Table 8 we present the frequency distribution of the cluster size for both bootstrap and asymptotic tests and for each[32] economic sector. Row totals provide an indication of the degree of convergence for each economic sector. Column totals provide information on the number of convergence clubs across sectors by cluster size. The asymptotic results are displayed in panel I and the bootstrap results are displayed in panel II. Overall, we observe a common pattern, namely a shift in the probability distribution towards a fewer number of clusters of larger size, and a commensurate increase in the extent of regional convergence. The total number of clusters for the asymptotic tests is 81, which falls by 32% to 55 clusters for the bootstrap test. This pattern is repeated for all sectors. Comparing column totals across the two tests is also informative since it gives the total number of clusters by cluster size, also shown in Figure 3. For the asymptotic test, more than 80% of the probability mass is distributed in clusters of size 4 or less, with approximately 10% of clusters of size 6 or more. In contrast, for the bootstrap test, approximately 50% of the clusters have a cluster size of 4 or less, with approximately 40% of clusters of size 6 or more.

Examining the results for each sector, for agriculture the size of the largest cluster generated by bootstrap critical values increases from seven to ten regions, with a commensurate decrease in the number of clusters of size 5 or less. Similarly for the manufacturing sector we observe an increase in the size of the largest cluster from six to nine regions and a decrease in the number of clusters of size 4 or less. In the market-service sector there is a reduction in the size of the largest cluster from nine to eight and for non-market services there is no change in the size of the largest cluster, but a substantial increase in clustering at the medium scale. In both service sectors there is a decrease in the number of clusters of size 4 or less.

Cluster Composition In establishing whether the composition of the clusters (i.e. the constituent regions) is changing between the two tests, we first collect the asymptotic (A) generated cluster outcomes in a $N \times N$ matrix $\mathbf{M}^A = \{m_{ij}^A\}$; element m_{ij}^A equals to 1 if regions i and j belong to the same cluster and zero otherwise. $\mathbf{M}^B = \{m_{ij}^B\}$ denotes the same for the bootstrap (B) generated cluster outcomes. The correlation parameter between the asymptotic, $\mathbf{M}^A$, and the bootstrap cluster

---

[32] In order to directly compare the bootstrap and asymptotic results in Corrado et al. (2005) we set $p_{\min}$ to be equal to 0.01 and the bandwidth $L = 2$. The number of bootstrap samples is set at 200.

pattern, $\mathbf{M}^B$, is then given by

$$\zeta_l = \left(\frac{\sum_{i=1}^{N}\sum_{j\neq i}^{N} m_{ij}^B \times m_{ij}^A}{\left(\sum_{i=1}^{N}\sum_{j\neq i}^{N} m_{ij}^B\right)^{1/2} \left(\sum_{i=1}^{N}\sum_{j\neq i}^{N} m_{ij}^A\right)^{1/2}} \right)^{1/2}, \tag{19}$$

where l indexes the set {Agriculture, Manufacturing, Market Services, Non-Market Services}. The results are reported in panel III of Table 8. With correlation coefficients ranging between 50% for manufacturing and 67% for agriculture, we note further evidence of a significant difference in the composition of the clusters generated by the asymptotic and bootstrap tests.[33]

Mean Income In order to assess the properties of each cluster we compute mean log per-capita income,[34] $\bar{y}_g$ for each test. The top panel of Figure 4 shows that the asymptotic test generates a distribution with a large number of small clubs while in the bootstrap test there are a fewer number of clusters of larger size. A visual impression of the oversized property of the asymptotic test of convergence is also evident in the distribution of the cluster mean of log per-capita income and in a relatively higher right kurtosis of this distribution, as presented in the lower panel of Figure 4. In this case an overrejection of the convergent null generates a distribution with a large number of small clubs characterised by a higher mean log per-capita income which results in a widening of the gap between the poorest and the richest clusters. In examining the comparable bootstrap distribution we observe a marked decrease in right kurtosis and a commensurate narrowing of the gap between the richest and the poorest cluster. Summary statistics are provided in the last three columns of panels I and II of Table 8. Note that for the bootstrap distribution the reduction in the gap between the richest and the poorest clusters is evident in a lower standard deviation of mean cluster per-capita income (from 15.2 to 5.4). The narrowing of the gap between the richest and poorest cluster translates into an increase in mean log per-capita income of the *poorest* cluster, $\bar{y}_{min}$, by around 24% (from 9.4 to 11.7) and a decrease in mean log per-capita income of the *richest* cluster, $\bar{y}_{max}$, by almost 50% (from 103 to 62.6). These results demonstrate the importance of the correct identification of convergence clubs. Given that many policy instruments are designed

---

[33] The method used in this paper to locate convergence clubs bypasses the particular problem of exactly how to utilize conditioning information in the model specification. Corrado and Weeks (2011) provide further information on how to interpret the results by confronting the resulting cluster composition, for both the asymptotic and the bootstrap tests, with a set of hypothetical clusters based on different theories and models of regional growth and development.

[34] Mean income is the cluster mean of log per-capita GVA.

to reduce the gap between the richest and the poorest regions, basing inference and policy decisions on the results of the asymptotic test would indicate the need for a stronger action than is actually needed when looking at the bootstrap test outcomes.

6 Conclusions

This study represents an extension of the multivariate test of stationarity which allows for endogenous identification of the number and composition of regional convergence clusters using sequential pairwise tests for stationarity. The main drawback of this approach is the short time-horizon which affects the size of the test. Our proposed bootstrap based extension to the sequential pairwise multivariate tests for stationarity performs well in Monte Carlo simulations in identifying and detecting correctly cluster membership when compared with the asymptotic version of the Hobijn and Franses (2000) approach. Based upon Monte Carlo evidence comparing the performance of CW with HF varying the number of countries, data span, club size and degree of persistence, indicate that detection rates of club membership (net of misclassifications) improve considerably when we implement the bootstrap. In operationalizing a bootstrap test of multivariate stationarity our results confirm the oversized property of the asymptotic test, and reveal a significantly greater degree of convergence. This evidence is gathered using both cross-country and regional data for the European Union for a number of industrial sectors. Our results show that by resolving the size distortion which afflicts the asymptotic test we find considerably more evidence of convergence in both the applications considered.

References

[1] Acemoglu, D. and Zilibotti, F. (1997). 'Was Prometheus Unbounded by Chance? Risk, Diversification and Growth', *Journal of Political Economy*, 105(7), pp. 710-751.

[2] Barro, R. J. and Sala-i-Martin, X. (1997). 'Technological Diffusion, Convergence and Growth', *Journal of Economic Growth*, 2(1), pp. 1-26.

[3] Basu, S. and Weil, D.N. (1998). 'Appropriate Technology and Growth', *Quarterly Journal of Economics*, 113, pp. 1025-1054.

[4] Baumol, W.J. (1986). 'Productivity Growth, Convergence and Welfare: What are the Long-Run Data Show', *The American Economic Review*, 76, pp. 1072-1085.

[5] Bernard, A. B. and Durlauf, S. N. (1995). 'Convergence in International Output', *Journal of Applied Econometrics*, 10(2), pp. 97-108.

[6] Bernard, A. B. and Durlauf, S. N. (1996).'Interpreting Tests of the Convergence Hypothesis', *Journal of Econometrics*, 71(1-2), pp. 161-73.

[7] Boldrin, M. and Canova, F. (2001). 'Inequality and Convergence: Reconsidering European Regional Policies', *Economic Policy*, 16(32), pp. 207-53.

[8] Breuer, Boucher J. , McNown, R., and Wallace, M. (1999). 'Series-specific Tests for a Unit Root in a Panel Setting with an Application to Real Exchange Rates', Discussion Paper No. 99-9, University of Colorado, June.

[9] Breitung, J. and M.H. Pesaran (2008). 'Unit Roots and Cointegration in Panels', by L. Matyas, and P. Sevestre (eds), *The Econometrics of Panel Data* (Third Edition), Ch. 9, pp. 279-322.

[10] Beylunioğlu, F., Stengos, T. and Yazgan, E. (2018), 'Detecting Convergence Clubs', *Macroeconomic Dynamics, forthcoming,* doi:10.1017/s1365100518000391

[11] Caner, M. and Kilian, L. (2001), 'Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate', *Journal of International Money and Finance,* 20, pp. 639-657.

[12] Carlino, G. A. and Mills, L. (1993). 'Convergence and US States: A Time Series Analysis', *Journal of Monetary Economics,* 32, pp. 335-46.

[13] Chatterji, M. (1992). ‘Convergence Clubs and Endogenous Growth’, *Oxford Review of Economic Policy*, 8(4), pp. 57-69.

[14] Cheshire, P. C. and Magrini, S. (2000). ‘Endogenous Processes in European Regional Growth: Convergence and Policy’, *Growth and Change*, 31, pp. 455-479.

[15] Ciccone, A. and Matsuyama, K. (1996). ‘Start-up Costs and Pecuniary Externalities as Barriers to Economic Development’, *Journal of Development Econonomics*, 49, pp. 33-59.

[16] Corrado, L., Martin, R. and Weeks, M. (2005). ‘Identifying and Interpreting Regional Convergence Clusters across Europe’, *The Economic Journal*, 115(502), pp. C133-C160.

[17] Corrado, L. and Weeks, M. (2011). ‘Tests of Convergence Clubs’, Cambridge Working Paper Series CWPE1031.

[18] Durlauf, S. N. and Johnson, P. (1995). ‘Multiple Regimes and Cross-Country Growth Behavior’, *Journal of Applied Econometrics*, 10(4), pp. 365-84.

[19] Durlauf, S. N., Johnson, P. A., and Temple, J. R. W. (2005). ‘Growth Econometrics’. In P. Aghion and S. N. Durlauf (eds.) *Handbook of Economic Growth*, Vol. 1A, North-Holland: Amsterdam, pp. 555-677.

[20] Efron, B. and Tibshirani, R. (1986). ‘Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy’, *Statistical Science*,1(1), pp. 54-75.

[21] Evans, P. (1998). ‘Using Panel Data to Evaluate Growth Theories,’ *International Economic Review*, 39 (1998), pp. 295-306.

[22] Fleissig, A. and Strauss, J. (2001). ‘Panel Unit-Root Tests of OECD Stochastic Convergence’, *Review of International Economics*, 9(1), pp. 153-162.

[23] Flores, R., P. Jorion, P.Y. Preumont and Szarfarz, A. (1999). ‘Multivariate Unit Root Tests of the PPP Hypothesis’, *Journal of Empirical Finance*, 6, pp. 335–353.

[24] Fujita, M., Krugman, P. and Venables, A. J. (1999). *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA: MIT Press.

[25] Fujita, M. and Thisse, J. F. (2002). *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth*, Cambridge: Cambridge University Press.

[26] Grossman, G. M. and Helpman, E. (1994). ‘Endogenous Innovation in the Theory of Growth’, *Journal of Economic Perspectives*, Vol. 8(1), pp. 23-44.

[27] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

[28] Harvey, A. and J. Bernstein, (2003). ‘Measurement and Testing of Inequality from Time Series of Deciles with an Application to U.S. Wages’, *The Review of Economics and Statistics*, MIT Press, 85(1), pp. 141-152.

[29] Hobijn, B. and Franses, P. H. (2000). ‘Asymptotically Perfect and Relative Convergence of Productivity’, *Journal of Applied Econometrics*, 15(1), pp. 59-81.

[30] Horowitz, L. (2001). ‘The Bootstrap’, in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, pp. 3159-3228, Amsterdam, Elsevier.

[31] Im, K.S., Pesaran, M. H. and Shin, Y. (2003). ‘Testing for Unit Roots in Heterogeneous Panels’, *Journal of Econometrics*, 115(1), pp. 53-74.

[32] Kapetanios, G. (2003). ‘Determining the Stationarity Properties of Individual Series in Panel Dataset’, Queen Mary University Department of Economics Working Paper No. 495.

[33] Kuo, B. and Tsong, C. (2005). ‘Bootstrap Inference for Stationarity’, Helsinki Center of Economic Research Discussion Paper No 50.

[34] Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). ‘Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. How Sure are we that Economic Time Series have a Unit Root?,’ *Journal of Econometrics*, 54(1-3), pp. 159-78.

[35] Levin, A., C. Lin, and C.J. Chu (2002), ‘Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties,’ *Journal of Econometrics*, 108, 1-24.

[36] Leybourne, S. and McCabe, B. (1994). ‘A Consistent Test for a Unit Root’, J*ournal of Business and Economic Statistics,* 12, pp. 157-166.

[37] Louveaux, F., Thisse, J. F. and Beguin, H. (1982). ‘Location Theory and Transportation Costs’, *Regional Science and Urban Economics*, 12(4), pp. 529-45.

[38] Lucas, R. (1988). ‘On the Mechanics of Economic Development’, *Journal of Monetary Economics*, 22(1), pp. 3-42.

[39] Maasoumi, E. and Wang, L. (2008), ‘Economic Reform, Growth and Convergence in China’, *The Econometrics Journal*, 11(1), pp. 128-154.

[40] Maddala, G.S. and Li, H. (1997). ‘Bootstrapping Cointegrating Regressions’, J*ournal of Econometrics*, 80, pp. 297-318.

[41] Maddala, G.S. and Wu, S. (1999). ‘A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test’, *Oxford Bulletin of Economics and Statistics*, 61, pp. 631–652.

[42] Moon, H.R. and Perron, B. (2007), ‘Efficient Estimation of the SUR Cointegration Regression Model and Testing for Purchasing Power Parity’, *Econometric Reviews*, 23, pp. 293-323.

[43] Murtagh, F. (1985). *Multidimensional Clustering Algorithms*, Physica-Verlag, Wurzburg.

[44] Pesaran, M. and A. Timmermann (1992). ‘A simple nonparametric test of predictive performance’, *Journal of Business and Economic Statistics* 10, 461-465.

[45] Pesaran, M. H. (2007). ‘A Pairwise Approach to Testing for Output and Growth Convergence’, *Journal of Econometrics*, Elsevier, 138(1), pp. 312-355.

[46] Phillips, P. and Sul, D. (2007) ‘Transition modeling and econometric convergence tests’ *Econometrica*, 75(6), 1771-1855.

[47] Quah, D.T. (1997). ‘Empirics for Growth and Distribution: Stratification, Polarization and Convergence Clubs’, *Journal of Economic Growth,* Vol. 2, pp. 27–59.

[48] Romer, P. (1986). ‘Increasing Returns and Long-Run Growth’, *Journal of Political Economy*, 94(5), pp. 1002-37.

[49] Solow, R. (1956). ‘A Contribution to the Theory of Economic Growth’, *Quarterly Journal of Economics,* 70(1), pp. 65–94.

[50] Swan, T.W. (1956), ‘Economic Growth and Capital Accumulation’, *Economic Record,* 32(63), pp. 334–361.

[51] Tan, C. M. (2009). ‘No One True Path: Uncovering the Interplay between Geography, Institutions, and Fractionalization in Economic Development’, *Journal of Applied Econometrics*, forthcoming.

[52] Taylor, M. P. and Sarno, L. (1998). ‘The Behavior of Real Exchange Rates during the Post-Bretton Woods Period,’ *Journal of International Economics*, 46(2), pp. 281-312.

[53] Weeks, M. and Yao Y. J. (2003). ‘Provincial Conditional Income Convergence in China, 1953–1997: A Panel Data Approach,’ *Econometric Reviews*, 22(1), pp. 59-77.

Table 1: Single Clubs Results

DataType				KS						PT					
N	m	ρ	T	HF			CW			HF			CW		
				0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
10	3	0.2	50	0.60	0.63	0.66	0.85	0.87	0.89	1.67	1.91	2.02	2.49	2.51	2.72
			100	0.73	0.78	0.83	0.76	0.79	0.81	1.96	2.16	2.34	2.26	2.28	2.49
		0.6	50	0.69	0.71	0.74	0.87	0.88	0.91	2.04	2.17	2.40	2.51	2.58	2.64
			100	0.73	0.81	0.82	0.77	0.81	0.81	2.02	2.42	2.52	2.29	2.33	2.29
	5	0.2	50	0.61	0.61	0.64	0.87	0.88	0.87	1.68	2.00	1.95	2.68	2.73	2.69
			100	0.69	0.76	0.73	0.79	0.82	0.82	2.10	2.18	2.31	2.62	2.68	2.69
		0.6	50	0.67	0.67	0.67	0.89	0.89	0.88	2.15	2.25	2.21	2.75	2.76	2.73
			100	0.71	0.76	0.77	0.80	0.83	0.84	2.21	2.40	2.40	2.62	2.63	2.69
20	3	0.2	50	0.47	0.47	0.57	0.66	0.74	0.73	1.60	2.18	2.43	2.87	3.08	3.53
			100	0.60	0.58	0.63	0.54	0.61	0.65	2.39	2.46	2.71	2.09	2.39	2.98
		0.6	50	0.58	0.60	0.64	0.72	0.81	0.81	1.82	1.98	2.35	3.14	3.46	3.51
			100	0.65	0.69	0.73	0.62	0.72	0.72	2.64	2.83	2.91	2.13	2.57	3.21
	5	0.2	50	0.60	0.58	0.58	0.87	0.88	0.87	2.74	2.88	2.94	3.84	3.86	3.82
			100	0.67	0.71	0.71	0.80	0.82	0.83	2.92	2.99	3.04	3.40	3.56	3.75
		0.6	50	0.63	0.63	0.66	0.92	0.91	0.90	2.34	2.56	2.55	3.91	3.89	3.72
			100	0.71	0.70	0.71	0.81	0.82	0.84	3.05	3.39	3.17	3.48	3.37	3.62
	7	0.2	50	0.61	0.60	0.58	0.90	0.87	0.86	2.78	2.80	2.74	3.98	3.79	3.79
			100	0.64	0.66	0.66	0.88	0.88	0.88	3.03	2.90	3.09	4.03	4.01	4.11
		0.6	50	0.63	0.61	0.62	0.91	0.88	0.86	2.13	2.67	2.68	4.06	3.90	3.83
			100	0.68	0.69	0.68	0.88	0.88	0.89	3.22	3.08	3.29	3.91	3.85	3.94
	10	0.2	50	0.52	0.48	0.47	0.73	0.72	0.70	2.42	2.48	2.37	3.28	3.25	3.24
			100	0.61	0.64	0.64	0.81	0.79	0.79	3.09	3.06	3.01	3.83	3.70	3.64
		0.6	50	0.56	0.50	0.52	0.76	0.74	0.74	2.24	2.18	2.28	3.36	3.33	3.36
			100	0.65	0.62	0.62	0.82	0.80	0.79	3.08	3.09	3.02	3.77	3.62	3.61
30	3	0.2	50	0.54	0.49	0.56	0.76	0.80	0.78	1.76	1.85	1.97	2.56	2.65	2.71
			100	0.59	0.62	0.65	0.72	0.76	0.77	2.11	2.33	2.56	2.21	2.31	2.38
		0.6	50	0.60	0.56	0.61	0.80	0.84	0.81	2.01	2.13	2.23	2.62	2.69	2.77
			100	0.65	0.67	0.72	0.73	0.76	0.77	2.13	2.43	2.55	2.25	2.38	2.37
	5	0.2	50	0.54	0.52	0.50	0.81	0.82	0.78	1.93	1.93	2.03	2.77	2.77	2.75
			100	0.59	0.61	0.63	0.76	0.77	0.76	2.22	2.40	2.31	2.55	2.62	2.62
		0.6	50	0.56	0.54	0.57	0.85	0.85	0.85	2.13	2.13	2.14	2.83	2.83	2.79
			100	0.64	0.67	0.66	0.79	0.79	0.80	2.29	2.40	2.43	2.57	2.66	2.67
	7	0.2	50	0.48	0.46	0.44	0.77	0.76	0.73	1.67	1.71	2.16	2.62	2.98	2.97
			100	0.60	0.57	0.57	0.82	0.82	0.81	2.06	2.09	2.29	1.94	2.20	2.35
		0.6	50	0.52	0.53	0.54	0.84	0.82	0.80	2.09	2.21	2.41	2.94	3.32	3.32
			100	0.63	0.64	0.62	0.85	0.85	0.85	2.22	2.49	2.75	2.24	2.62	2.65
	10	0.2	50	0.43	0.41	0.41	0.67	0.67	0.65	2.46	2.44	2.49	3.76	3.82	3.80
			100	0.53	0.52	0.53	0.74	0.73	0.73	2.68	2.94	3.00	3.24	3.34	3.40
		0.6	50	0.49	0.42	0.43	0.72	0.71	0.69	2.63	2.67	2.89	3.98	3.98	3.94
			100	0.61	0.60	0.60	0.78	0.77	0.76	2.85	2.86	2.97	3.31	3.37	3.48

Table 2: Multiple Club Results

Data Type										
N	m	k	ρ	T	HF			CW		
					0.01	0.05	0.1	0.01	0.05	0.1
10	4,4	2	0.2	50	34.20%	28.00%	17.00%	28.00%	22.00%	14.40%
				100	51.80%	39.00%	33.00%	61.20%	45.00%	40.00%
			0.6	50	38.80%	26.60%	19.00%	35.00%	24.00%	17.00%
				100	44.00%	38.00%	34.80%	54.00%	45.00%	37.00%
	3,3,2	3	0.2	50	1.20%	1.00%	1.00%	15.60%	15.60%	14.40%
				100	10.40%	8.20%	5.60%	27.00%	25.00%	22.00%
			0.6	50	13.00%	14.00%	10.00%	13.00%	15.60%	14.40%
				100	3.20%	2.60%	1.00%	20.00%	17.80%	15.00%

Table 3: List of Countries (PWT)

	Country		Country		Country
AGO	Angola	GIN	Guinea	NLD	Netherlands
ARG	Argentina	GMB	Gambia	NOR	Norway
AUS	Australia	GNB	Guinea Bissau	NZL	New Zealand
AUT	Austria	GRC	Greece	PAK	Pakistan
BDI	Burundi	GTM	Guatemala	PAN	Panama
BEL	Belgium	GUY	Guyana	PER	Peru
BEN	Benin	HKG	Hong Kong	PHL	Phillipines
BGD	Bangladesh	HND	Honduras	PNG	Papua N. Guinea
BOL	Bolivia	HTI	Haiti	PRI	Puerto Rico
BRA	Brazil	HVO	Burkina Faso	PRT	Portugal
BRB	Barbados	IDN	Indonesia	PRY	Paraguay
BUR	Myanmar	IND	India	RWA	Rwanda
BWA	Botswana	IRL	Ireland	SEN	Senegal
CAF	Central African Rep.	IRN	Iran	SGP	Singapore
CAN	Canada	ISL	Iceland	SLV	El Salvador
CHE	Switzerland	ISR	Israel	SOM	Somalia
CHL	Chile	ITA	Italy	SUR	Suriname
CIV	Ivory Coast	JAM	Jamaica	SWE	Sweden
CMR	Cameroon	JOR	Jordan	SWZ	Swaziland
COG	Congo	JPN	Japan	SYC	Seychelles
COL	Colombia	KEN	Kenya	SYR	Syria
CPV	Cape Verde Is.	KOR	Korea	TCD	Tcad
CRI	Costa Rica	LKA	Sri Lanka	TGO	Togo
CSK	Czechoslovakia	LSO	Lesotho	THA	Thailand
CYP	Cyprus	LUX	Luxembourg	TTO	Trinidad/Tobago
DEU	West Germany	MAR	Morocco	TUN	Tunisia
DNK	Denmark	MDG	Madagascar	TUR	Turkey
DOM	Dominican Rep.	MEX	Mexico	UGA	Uganda
DZA	Algeria	MLI	Mali	URY	Uruguay
ECU	Ecuador	MLT	Malta	USA	United States
EGY	Egypt	MOZ	Mozambique	VEN	Venezuela
ESP	Spain	MRT	Mauritania	YUG	Former Yugoslavia
FIN	Finland	MUS	Mauritius	ZAF	South Africa
FJI	Fiji	MWI	Malawi	ZAR	Zaire
FRA	France	MYS	Malaysia	ZMB	Zambia
GAB	Gabon	NAM	Namibia	ZWE	Zimbabwe
GBR	United Kingdom	NER	Niger		
GHA	Ghana	NGA	Nigeria		

Table 4: Joint Frequency Distribution (PWT)

I: Asymptotic
Number of Clusters

Cluster size	1	2	3	4	5	6	7	Total Clusters
Perfect	29	22	9	3	0	0	0	63
Relative	2	21	12	4	2	1	0	42

II: Bootstrap
Number of Clusters

Cluster size	1	2	3	4	5	6	7	Total Clusters
Perfect	1	4	4	8	3	5	2	27
Relative	1	3	4	7	3	6	2	26

Table 5: Asymptotic: Relative Convergence (PWT)

No	Countries					
1	AUS	DNK	KEN	LUX	MUS	ZAF
2	AUT	ESP	ISR	ITA	PRI	
3	CAN	CSK	ECU	GRC	IRL	
4	BDI	HVO	MLI	MWI		
5	BGD	BUR	HND	NZL		
6	CAF	IND	NER	UGA		
7	GUY	JOR	SLV	SYC		
8	AGO	GHA	HTI			
9	BEN	GIN	VEN			
10	BOL	LKA	PNG			
11	BRB	IDN	THA			
12	CIV	COG	MAR			
13	CMR	CRI	NGA			
14	CPV	GNB	RWA			
15	FIN	ISL	TTO			
16	FJI	NAM	PER			
17	IRN	PRT	YUG			
18	MRT	PAK	SOM			
19	MYS	SWZ	TUR			
	Clusters with two countries					
20	ARG	GMB	21	BEL	NOR	
22	BRA	SUR	23	BWA	MLT	
24	CHE	USA	25	CHL	GAB	
26	COL	JAM	27	CYP	SGP	
28	DEU	FRA	29	DOM	SWE	
30	DZA	GTM	31	EGY	ZWE	
32	GBR	NLD	33	HKG	KOR	
34	LSO	TGO	35	MDG	ZMB	
36	MEX	URY	37	MOZ	SEN	
38	PAN	SYR	39	PRY	TUN	
40	TCD	ZAR				
	Two separate countries					
41	JPN					
42	PHL					

Listed according to size. $p_{\min} = 0.01$ and $l = 2$.

Table 6: Bootstrap: Relative Convergence (PWT)

No	Countries						
1	COL	JAM	MYS	NAM	PAN	SYR	TUR
2	AGO	BGD	CMR	GHA	MRT	PAK	SOM
3	CHE	CHL	CSK	GRC	MEX	USA	
4	GUY	JOR	PNG	SLV	SYC	TUN	
5	CRI	GAB	IRN	MUS	SUR	ZAF	
6	CIV	COG	LKA	MAR	PHL	THA	
7	AUT	FIN	ISL	ITA	JPN	TTO	
8	BDI	BUR	GIN	HVO	MLI	MWI	
9	LSO	RWA	TCD	TGO	ZAR		
10	CPV	GMB	GNB	NER	UGA		
11	DEU	DNK	FRA	LUX	NZL		
12	CYP	PRT	SGP	YUG			
13	DZA	ECU	GTM	SWZ			
14	BOL	DOM	KOR	PRY			
15	BRA	FJI	MLT	PER			
16	CAF	IND	KEN	NGA			
17	BEL	GBR	NLD	NOR			
18	BEN	MDG	ZMB	ZWE			
19	AUS	CAN	SWE				
20	BRB	HKG	IRL				
21	BWA	MOZ	SEN				
22	ESP	ISR	PRI				
	Clusters with two countries						
23	HTI	IDN					
24	EGY	HND					
25	ARG	URY					
	One separate country						
26	VEN						

Listed according to size. $p_{\min} = 0.01$ and $l = 2$.

The number of bootstrap samples is set at 200.

Table 7: NUTS1 code

Code	Country		Code	Country	
AT	*Austria*		IE	*Ireland*	
AT1		Ostosterreich			
AT2		Sudosterreich	IT	*Italy*	
AT3		Westosterreich	IT1		Nord Ovest
BE	*Belgium*		IT2		Lombardia
BE1		Region Bruxelles-Capital-Brussels	IT3		Nord Est
		Hoofdstedelijke Gewest	IT4		Emilia-Romagna
BE2		Vlaams Gewest	IT5		Centro
BE3		Region Wallonne	IT6		Lazio
DE	*Germany*		IT7		Abruzzo-Molise
DE1		Baden-Wurttemberg	IT8		Campania
DE2		Bayern	IT9		Sud
DE3		Berlin	ITA		Sicilia
DE5		Bremen	ITB		Sardegna
DE6		Hamburg	LU	*Luxembourg*	
DE7		Hessen			
DE9		Niedersachsen	NL	*Netherlands*	
DEA		Nordrhein-Westfalen	NL1		Noord-Nederland
DEB		Rheinland-Pfalz	NL2		Oost-Nederland
DEC		Saarland	NL3		West-Nederland
DEG		Thuringen	NL4		Zuid-Nederland
DK	*Denmark*		PT	*Portugal*	
			PT1		Continente
ES	*Spain*		SE	*Sweden*	
ES3		Comunidad de Madrid			
ES4		Centro	UK	*United Kingdom*	
ES5		Este	UKC		North East
ES6		Sur	UKD		North West
ES7		Canarias	UKE		Yorkshire and
F1	*Finland*				Humber
			UKF		East Midland
FR	*France*		UKG		West Midlands
FR1		Ile de France	UKH		East of England
FR2		Bassin-Parisien	UK1		London
FR3		Nord Pas de Calais	UKJ		South East
FR4		Est	UKK		South West
FR5		Ouest	UKL		Wales
FR6		Sud-Ouest	UKM		Scotland
FR7		Centre-Est			
FR8		Mediterranee			
GR	*Greece*				
GR1		Voreia Ellada			
GR2		Kentriki Ellada			
GR3		Attiki			
GR4		Nisia Aigaiou, Kriti			

Table 8: Joint Frequency Distribution

I: Asymptotic
Number of Clusters

Summary Statistics

Cluster size	1	2	3	4	5	6	7	8	9	10	Total Clusters	$\sigma_{\bar{y}}$	$\bar{y}^{min}$	$\bar{y}^{max}$
Agriculture	0	3	7	2	4	1	1	0	0	0	18			
Manufacturing	0	7	9	4	1	1	0	0	0	0	22			
Market Service	1	9	3	6	0	0	1	0	1	0	21			
Non-market Service	1	6	7	2	1	1	1	1	0	0	20			
Total Clusters	2	25	26	14	6	3	3	1	1	0	81	15.2	9.4	103

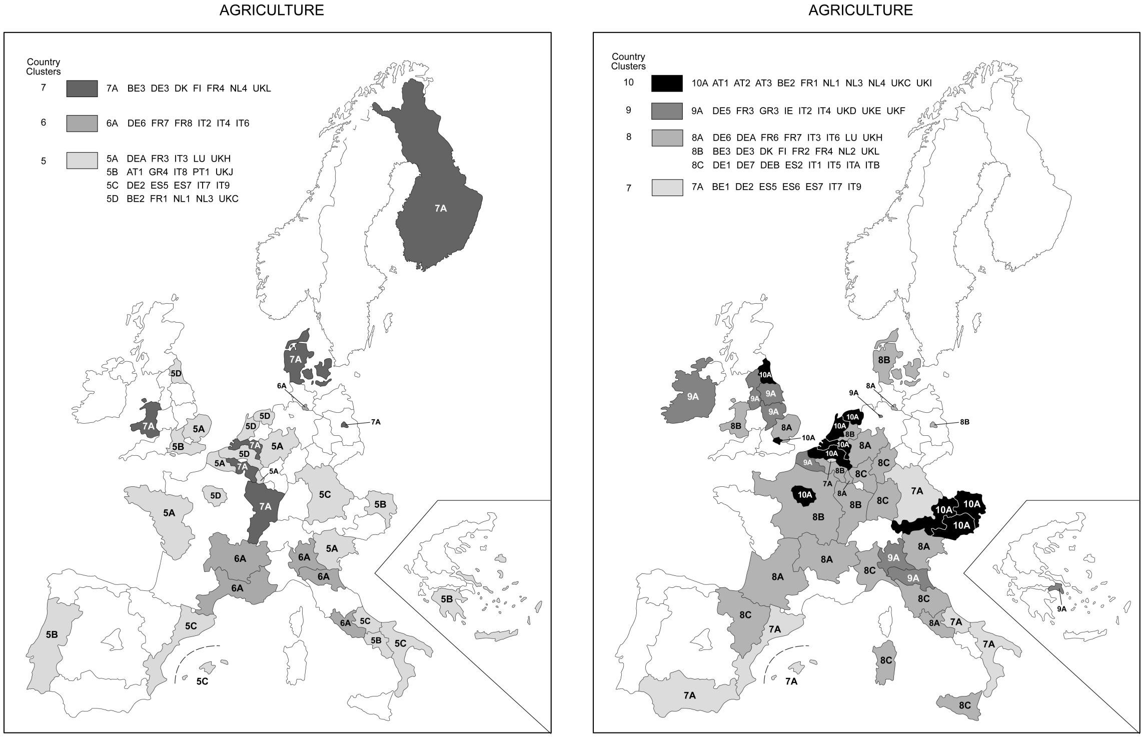
II: Bootstrap
Number of Clusters

Cluster size	1	2	3	4	5	6	7	8	9	10		$\sigma_{\bar{y}}$	$\bar{y}^{min}$	$\bar{y}^{max}$
Agriculture	0	3	1	1	1	0	1	3	1	1	12			
Manufacturing	0	2	5	1	2	3	0	1	1	0	15			
Market Services	0	1	3	2	2	4	1	1	0	0	14			
Non-market Services	0	1	3	2	3	3	0	2	0	0	14			
Total Clusters	0	7	12	6	8	10	3	8	2	2	55	5.4	11.7	62.6

III
Correlation Between Asymptotic and Bootstrap Cluster Outcomes

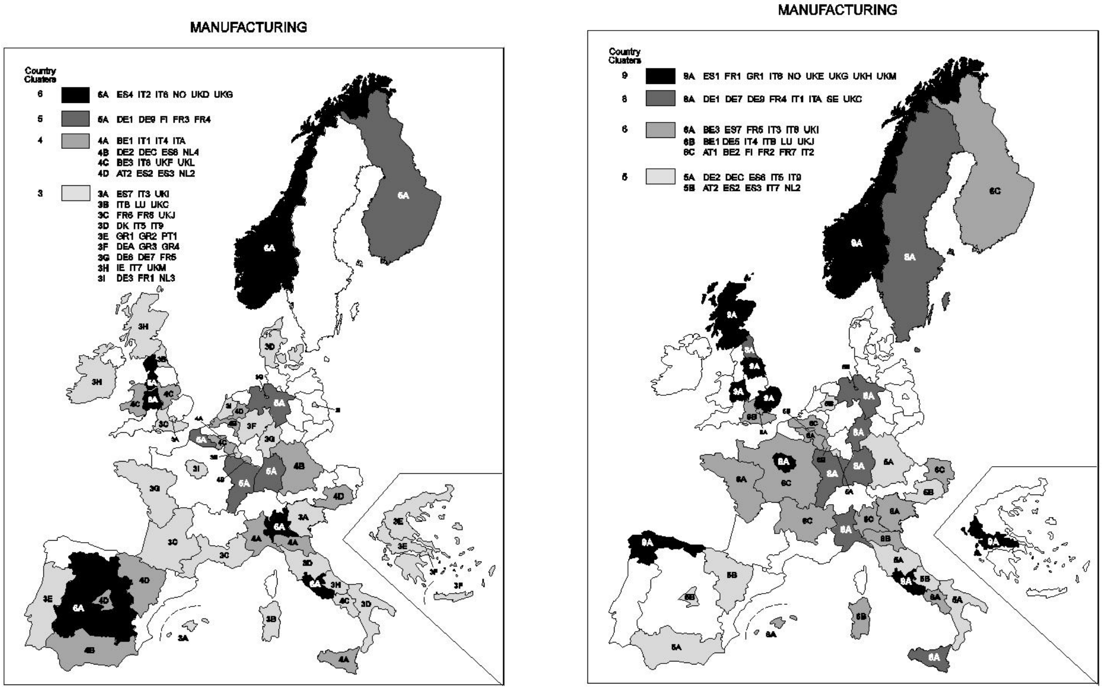
Agriculture	0.672
Manufacturing	0.509
Market Services	0.557
Non-Market Services	0.591

NB: $\sigma_{\bar{y}}$ denotes the standard deviation of cluster means. $\bar{y}^{min}$ and $\bar{y}^{max}$ denote the Min and Max of cluster means.



(a) Relative Convergence in Agriculture: Asymptotic Results

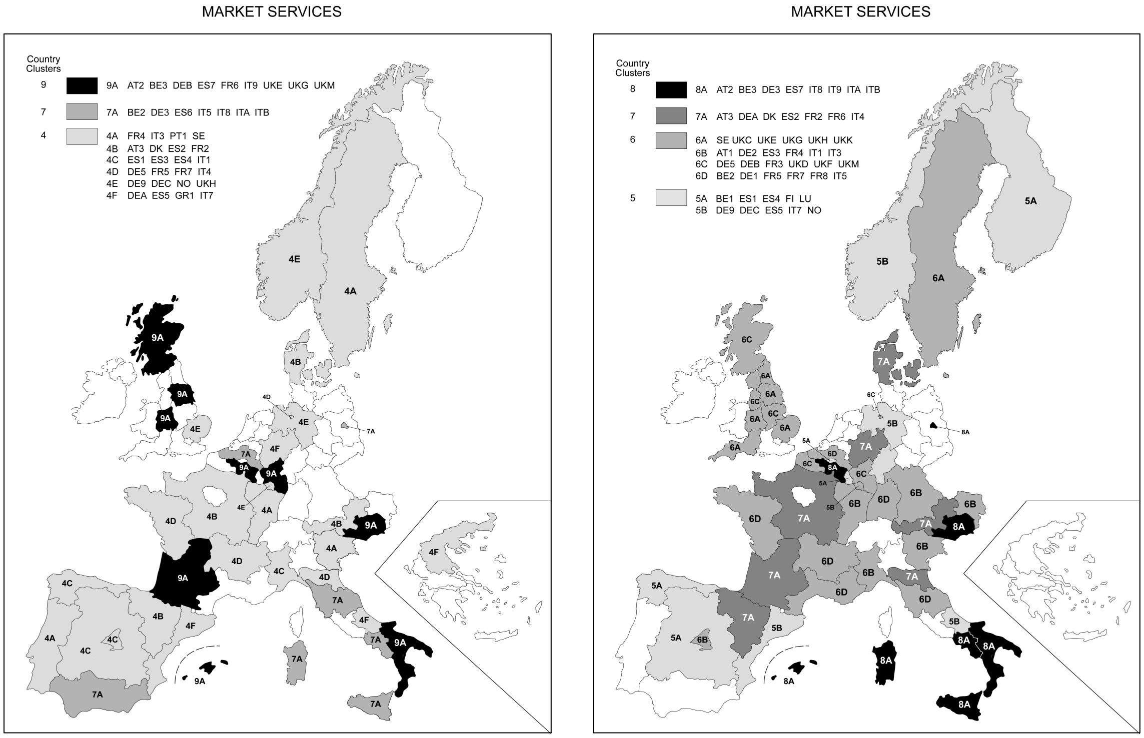
(b) Relative Convergence in Agriculture: Bootstrap Results



(c) Relative Convergence in Manufacturing: Asymptotic Results

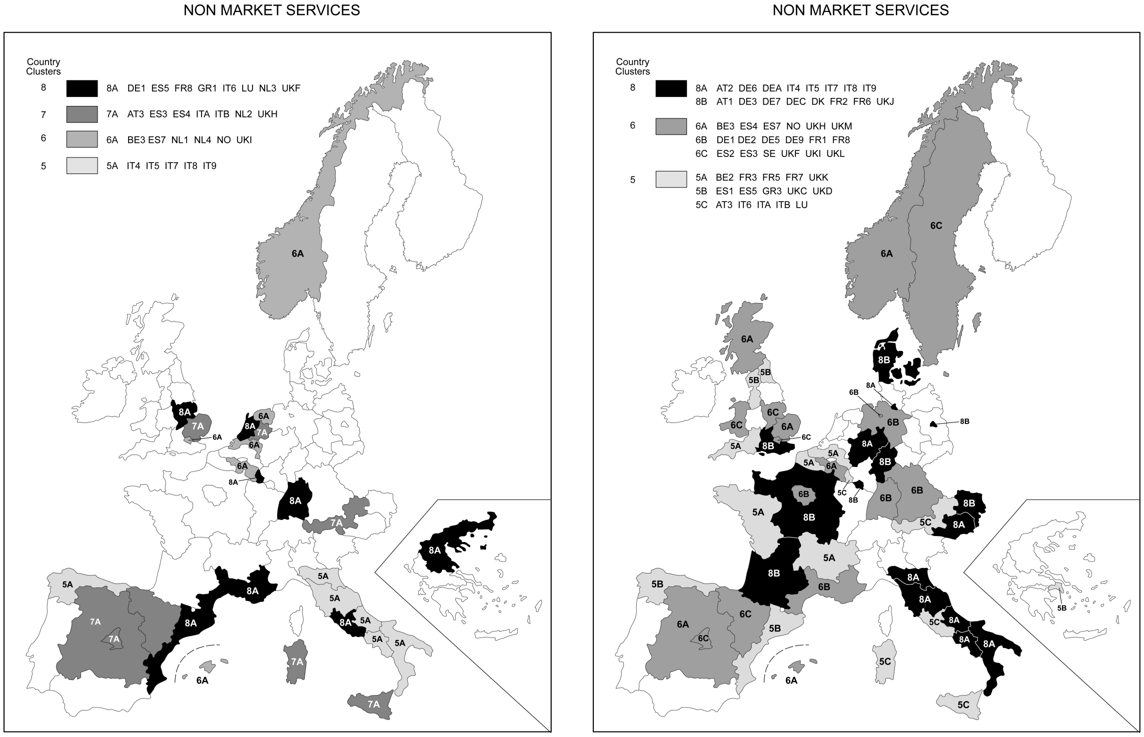
(d) Relative Convergence in Manufacturing: Bootstrap Results

Figure 1: Asymptotic and Bootstrap Results for Agriculture and Manufacturing



(a) Relative Convergence in Market Services: Asymptotic Results

(b) Relative Convergence in Market Services: Bootstrap Results

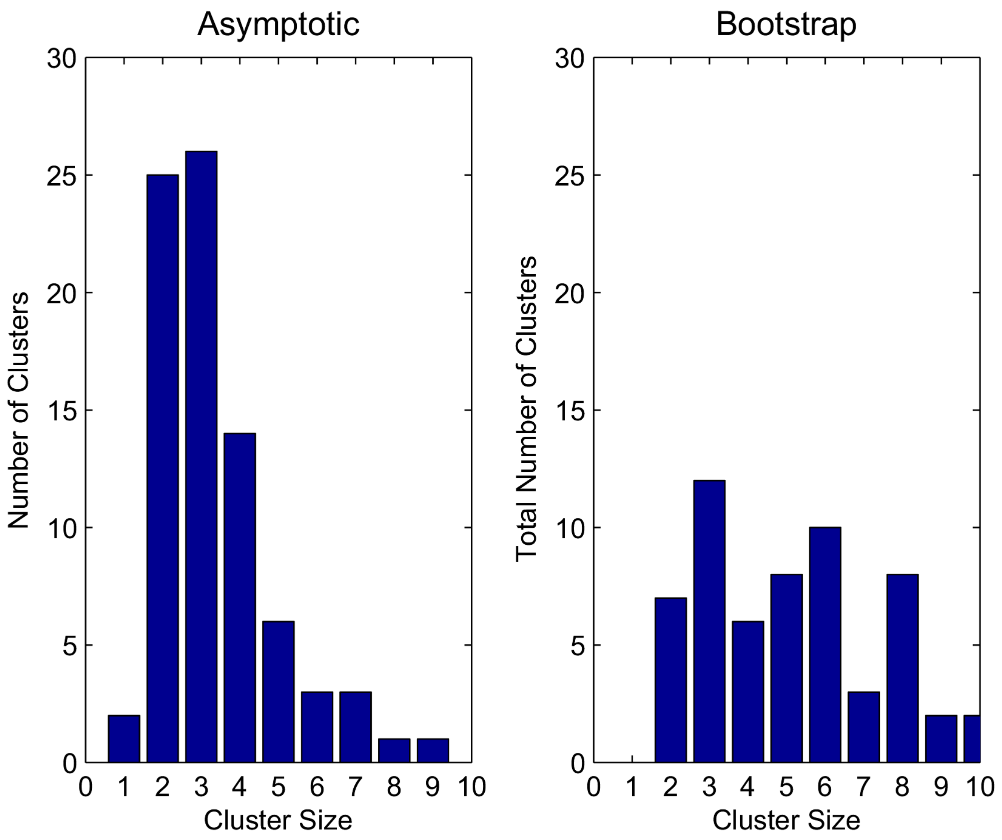


(c) Relative Convergence in Non-Market Services: Asymptotic Results

(d) Relative Convergence in Non-Market Services: Bootstrap Results

Figure 2: Asymptotic and Bootstrap Results for Non-Market and Market Services

Figure 3: The Distribution of Cluster Size.



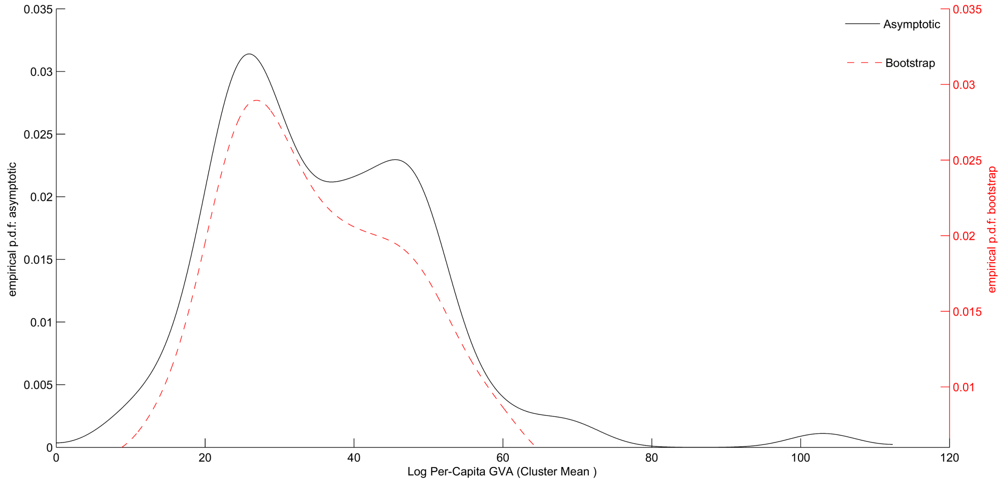


Figure 4: The distribution of average log per-capita GVA by cluster: All sectors.

Skewness (Asymptotic) 1.29 (Bootstrap) 0.27
Kurtosis (Asymptotic) 6.82 (Bootstrap) 2.20