# Tests for Convergence Clubs

Luisa Corrado and Melvyn Weeks

12 January 2011

CWPE 1110

# Tests for Convergence Clubs*

Luisa Corrado†

Faculty of Economics, University of Cambridge and University of Rome Tor Vergata

Melvyn Weeks‡

Faculty of Economics and Clare College, University of Cambridge

January 27, 2011

**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 tests which are founded upon an omnibus null hypothesis are characterised by a number of problems. In this paper we consider a broad class of tests of convergence based on multivariate time series and panel data methodologies, and track a gradual progression away from tests based on an omnibus null, to sequential tests and tests that are founded upon multiple pairwise comparisons. In a previous study Corrado, Martin and Weeks (2005) test for regional convergence across the European Union allowing for an endogenous selection of regional clusters using a multivariate test for stationarity. Given that the time series are relatively short, there are potential problems in basing inference on asymptotic results for stationarity tests. To circumvent this problem we bootstrap the stationarity test and explore the robustness of the cluster outcomes. In general our results show that the size distortion which afflicts the asymptotic tests, and resulting in a bias towards finding less convergence, is resolved when we apply the bootstrap generated critical values. To interpret the composition of the resulting convergence clusters, the latter are tested against a variety of possible groupings suggested by recent theories and hypotheses of regional growth and convergence.

**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.

†Faculty of Economics, University of Cambridge and University of Rome Tor Vergata. Luisa Corrado gratefully acknowledges the Marie-Curie Intra European fellowship 039326.

‡Faculty of Economics and Clare College, University of Cambridge, Cambridge CB3 9DD, UK. Email: mw217@econ.cam.ac.uk.

# 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 its is evident that the correct identification 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 any 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. The emergence over the past decade of New Economic Geography[2] 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 (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 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 validity of constructing such a large intersection null hypothesis is questionable in the first place. 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

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

[2] 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).

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 no subsequent 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.

This paper makes a number of contributions to the literature. First, we consider the pairwise approach of Hobijn and Franses (2000) in the light of a number of significant developments in tests designed to detect convergence. These developments are set against a backdrop where in many applications common in testing for convergence, the number of cross-sectional units is large, such that tests which are founded upon an omnibus null are not informative in the case of a rejection. Second, although many of the datasets used by analysts to detect convergence are of the large $N$ variant, in many instances $T$ is small. As a consequence 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 a recursive bootstrap test for stationarity which is designed to detect multiple convergence clubs without prespecification of group membership. We then compare the asymptotic and bootstrap generated cluster outcomes. By resolving the size distortion which afflicts the asymptotic test we find more evidence of convergence and also find a stronger correlation between the convergence patterns suggested by different theories of regional growth and the observed regional convergence clusters.

The paper is structured as follows. Section two presents a brief overview of the alternative methodologies to analyze convergence. Section three reviews existing tests for convergence clubs, and in section four we present the bootstrap version of the test. Section five describes the data. In section six we discuss our findings and a number of conclusions are offered in section seven.

## 2 Tests of Convergence

Tests of convergence are either framed around the null hypothesis of a unit root or around the null of stationarity. The different representations of both the null and alternative hypotheses reflect distinct economic concepts of convergence. For example, an analyst with country-level data might consider whether per capita output equalizes across industrialised and developing countries. With country-level data starting conditions are considered to be important and economies with lower levels of initial capital endowment are more likely to be far from their respective (or shared) steady-states. Here the notion of convergence has a statistical analogue in testing for unit roots in output differences. In contrast another analyst might consider the stability of output differences across one or more pairs of regions within a *single* industrialised country, and in this case testing for a stationary difference between per capita income is likely to be more appropriate.

In this section we review existing approaches to test for convergence. We focus on the form of the null and alternative hypothesis across a number of testing strategies and, related, the utility of such approaches as the number of regions increases. We start by examining the use of omnibus tests in the context of both multivariate time series and panel data approaches to testing for convergence. A critical problem with omnibus tests is that with large $N$ and $T$ panels the likelihood of rejecting the null increases with no obvious information on the form of the rejection. The *mixed* panel approach allows for some of the series to have a unit root while the rest to be stationary. Additional contributions have included the use of both sequential and pairwise testing strategies, as analysts have, in a number of instances, moved away from reliance of omnibus null tests of hypotheses to provide additional inference at a finer level. These tests 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.

## 2.1 Omnibus Tests

The use of time-series evidence to test for convergence was initiated by the seminal papers of Bernard and Durlauf (1995, 1996). A multi country definition of relative convergence asks whether the long-run forecast of all output differences with respect to a benchmark economy tend to a country-specific constant as the forecasting horizon tends to infinity[3]. We may then write

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

where $y_{(i1),t+s} = y_{it+} - y_{1t+}$, and $\mu_{1j}$ is a finite constant[4]. In an empirical application the authors test for output convergence across 15 OECD countries over the period 1900-1987, with output deviations for country $i$ relative to a US output benchmark given by $\Delta y_{(i1),t} = y_{(i1),t} - y_{(i1),t-1}$. The multivariate WOLD representation of the $N-1\times 1$ vector of output deviations $\Delta \mathbf{y}_{1,t} = \{\Delta y_{(i1),t}\}$ is given by

$$\Delta \mathbf{y}_{1,t} = \boldsymbol{\alpha}_1 + \sum_{l=1}^{\infty} \mathbf{C}(l)\boldsymbol{\varepsilon}_{1,t}, \tag{2}$$

where $\boldsymbol{\alpha}_1$ is a $(N-1)\times 1$ vector of intercepts, $\mathbf{C}(l)$ is a $(N-1)\times(N-1)$ matrix of polynomials in the lag operator $l$, and $\boldsymbol{\varepsilon}_{1,t} = \{\varepsilon_{(i1),t}\}$ is $(N-1)\times 1$ vector of error terms. Testing for non-convergence can be formalised as a condition on the rank of the $(N-1)\times(N-1)$ spectral density matrix of the innovation sequence $\boldsymbol{\Sigma}$. The main idea behind the test is that if multiple time series are cointegrated, the cointegrating relationships naturally reduces the variability in output deviations[5]. We can then use the moving average representation of the series in (2) to test

[3]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)).

[4]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.

[5]The system unit root test proposed by Bernard and Durlauf is based on the Phillips-Ouliaris test (1988).

for the number of linearly independent stochastic trends, with the null and alternative hypotheses given by

$$\begin{aligned} H_0 &: \quad rank(\mathbf{\Sigma}) = N-1 \\ H_1 &: \quad rank(\mathbf{\Sigma}) < N-1 \end{aligned}$$

For the joint null of non convergence *not* to be rejected there must exist $N$-1 distinct stochastic trends (unit roots) in output deviations. There exist a number of problems with the system unit-root test. First, the testing procedure is sensitive to the choice of the benchmark country.[6] 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 the convergence groups. Third, given the system properties of the test, a dimensionality constraints 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). However, as pointed out by Breitung and Pesaran (2008), panel data unit root tests have a number of shortcomings which are particularly pertinent in the context of testing for convergence. 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] 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. However, it is worth noting that the adoption of a multivariate ADF test while resolving the issue of cross-sectional dependence reintroduces the same dimensionality problem which characterises the multivariate unit-root test proposed by Bernard and Durlauf (1995).

In the context of testing for convergence a critical problem with panel unit root tests is that as $N$ becomes large the likelihood of rejecting the omnibus null increases with no information on the exact form of the rejection. The ADF regression to test for the convergence of output for region $i$ relative to a benchmark region 1 is given by

$$\begin{aligned} \Delta y_{(i1),t} &= \alpha_{(i1)} + \beta_{(i1)}t + \gamma_{(i1)}y_{(i1),t-1} + \sum_{k=1}^{p_i} \phi_{(i1),k}\Delta y_{(i1),t-k} + \varepsilon_{(i1),t} \\ i &= 2, ..., N \qquad t = 1, ..., T \end{aligned} \tag{3}$$

where $\alpha_{(i1)}$ is an intercept, $\beta_{(i1)}$ is the coefficient on the time trend, $\gamma_{(i1)}$ is the unit-root parameter,

[6] An alternative is the system cointegration test due to Johansen (1988) which is applied directly to the $N$ series of output levels (assumed to be $I(1)$),to determine the rank of the long-run cointegrating matrix which gives the number of cointegrating relationships. The omnibus null hypothesis of non-convergence is *not* rejected if the rank of the cointegrating matrix is less than $N$.

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

[8] Other notable example of second generation of panel unit root tests with cross-sectional dependence include Pesaran (2007) and Moon and Perron (2004).

and $\phi'_{(i1)}$ s are the autoregressive parameters.[9] $p_i$ is some sufficiently large integer, and $\varepsilon_{(i1),t}$ is a serially uncorrelated error term with a zero mean and a finite variance. The null and alternative hypotheses are given by

$$H_0 \quad : \quad \gamma_{(i1)} = 0 \qquad \forall i \neq 1, \tag{4}$$

$$H_1 \quad : \quad \gamma_{(i1)} < 0 \qquad \text{for at least one } i \neq 1 \tag{5}$$

A *mixed* alternative hypothesis allows, conditional on a rejection of the null, for some of the series to have unit roots. In this instance the alternative hypothesis has the following form

$$H_1^* : \begin{array}{ll} \gamma_{(i1)} < 0, & i = 2, ..., N_1 \\ \gamma_{(i1)} = 0, & i = N_1 + 1, ..., N, \end{array}$$

where $N_1$ denotes the number of stationary series. Given that $H_1^*$ has a mixed (heterogeneous) structure, the event of rejecting the null allows for a non zero fraction $\delta = N_1/N$ of the series to be stationary. However, as Breitung and Pesaran (2008) and Im et al. (2003) note, the test does not provide any guidance as to the magnitude of $\delta$, or the identity of the particular panel members for which the null hypothesis is rejected.

## 2.2 Sequential Tests

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, among others,[10] Kapetanios (2003, 2008). Specifically, Kapetanios employs a sequence of unit root tests of panels of decreasing size to separate stationary and nonstationary series[11], therefore allowing an endogenous identification of the number and identity of stationary series. Although the method is applied to series in levels, it is readily extended to output deviations with respect to a benchmark economy[12]. The starting point is the omnibus null hypothesis that all series have distinct unit root processes. If the null is not rejected then the procedure stops. If the null is rejected the series with the most evidence in favour of stationarity (i.e. with the minimum individual DF test statistic) is removed and the test is performed on the remaining series. This procedure iterates until the unit root null does not reject. At the point of termination, the result is the partition of the series into stationary ($S$) and nonstationary groups. In the case where the individual series are defined relative to a benchmark, the set of stationary deviations, namely $\mathbf{\Delta} y_{(i1),t} \ \ \forall i \in S$, represents a convergent group of countries.

Although a positive development there are a number of limitations of this approach. 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

[9] All roots of the polynomial $\sum_{k=1}^{p_i} \phi_{(i1),k} L^k$ lie outside the unit circle. $L$ is the lag operator.

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

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

[12] It is worth noting that although this sequential procedure proposed by Kapetanois is based upon a unit root null, as the author states, there is nothing preventing the use of such a procedure in conjunction with a different test statistic, such as KPSS.

classification of the $N$ series into two groups whereas there may be many more groups which are left unidentified. 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.

## 2.3 Pairwise Tests

A problem with the Kapetanois test when applied to output deviations 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 $m$ 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

$$\Pr(\cap_{i=1,...,N,j=i+1,...,N}|y_{it+s} - y_{jt+s}| < c|I_t) > \pi \tag{6}$$

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

For the $i, j^{th}$ country pair the unit root test is based upon the ADF regression

$$\Delta y_{(ij),t} = \alpha_{(ij)} + \beta_{(ij)}t + \gamma_{(ij)}y_{(ij),t-1} + \sum_{k=1}^{p_i} \phi_{(ij),k}\Delta y_{(ij),t-k} + \varepsilon_{(ij),t} \qquad t = 1, ..., T, \tag{7}$$

where all parameters are defined analagously to those in (3). For the stationary null, Pesaran utilises the Kwiatkowski *et al.* (1992) test (hereafter KPSS). 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

$$\widehat{\varepsilon}_{(ij),t} = y_{(ij),t} - \widehat{\alpha}_{(ij)} - \widehat{\beta}_{(ij)}t. \tag{8}$$

Defining $h_t = y_{(ij),t} - \frac{1}{T}\sum_{t=1}^{T} y_{(ij),t}$, the test statistic for level stationarity is given by

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

where $\bar{S}_t = \sum_{s=1}^{t} h_s$ denotes a partial sum process and

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}^2_{(ij),t} + 2\frac{1}{T}\sum_{k=1}^{L}\omega(k, L)\sum_{t=k+1}^{T}\widehat{\varepsilon}_{(ij),t}\widehat{\varepsilon}_{(ij),t-k}, \tag{9}$$

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. Pesaran finds that the proportion of output gaps for which the unit root and stationary null is rejected is close to the significance level, such that there is very little evidence for convergence.

We note that 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. Further, since only a single null hypothesis is tested there is no need for multiplicity corrections. However, although the pairwise tests of convergence proposed by Pesaran (2006) 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. In a recent study Moon and Peron (2009) focus on this question, utilising a multiple testing strategy to determine the stationarity properties of *individual* series in a panel using the false discovery rate (FDR)[13]. Whereas multiplicity corrections, such as those based on FDR,[14] are designed to control the false discovery rate across a series of $N$ *simultaneous* tests, and as a consequence facilitate inference at the level of individual comparisons, we require a testing procedure that can identify $J \leq N$ groups. Below we examine various testing strategies for club convergence and in particular the sequential testing procedure proposed by Hobijn and Franses (2000).

# 3 Tests for Convergence Clubs

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). In a number of studies the identification of convergence clubs has been achieved exogenously through testing in conjunction with a pre-classification of clubs using parametric techniques. 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 Massoumi 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. Durlauf and Johnson (1995) and more recently Tan (2009) utilize a regression tree approach which again utilises exogenous information in the form of conditioning variables.

In contrast, non-parametric techniques place fewer restrictions on the identification of groups and have generally been used in situations where both the determinants of growth and the number and composition of convergence clubs are unknown. For example, Quah (1997) has studied convergence by analyzing the empirical distribution of per-capita income across economies at

[13] The control of FDR is based upon two desirable characteristics of a multiple testing regime: (i) the number of false positives is important, (ii) as the number of comparisons increases the potential cost of any one erroneous rejection is likely to fall.

[14] The multi-step procedure to control FDR, proposed by Benjamini and Hochberg (1995), collects p-values in the $N \times 1$ rank ordered vector $\mathbf{p}^o = \{p_{(i)}\}$, the ordering reflecting increasing evidence in support of the null. Inference proceeds by adjusting critical values dependent upon the relative magnitude of the $p$ values. Starting with the least evidence *against* the null, $p_{(N)}$ is compared with critical value $\alpha$. As evidence for the null falls the threshold critical value is adjusted downwards. Significant hypotheses are selected given $\hat{k} = \mathtt{max}\{1 \leq k \leq N : p_k \leq \frac{\alpha k}{N}\}$.

different time periods. The so called distribution dynamics takes place when individual economies transit from one part of the distribution to another across time and allows the revelation of possible bi-modal twin-peaked distribution which is discretised in different income categories. As a result, both the number of groups and their composition is identified exogenously on the basis of the empirical distribution of per-capita income in each time period. Canova (2004) employs a predictive density (marginal likelihood) approach, treating the number of groups and composition thereof as unknown parameters. A number of factors that may explain the formation of convergence clubs are specified a priori: the most likely determinant of the groupings is that factor which maximizes the predictive power of the model. For example, in considering the importance of location regions are ordered on the basis of geographical proximity. As Canova notes, a problem with this approach is that for regional data there are few usable indicators on which to order units. In addition since the approach is heavily parametric, limits on the maximum number of clusters are imposed.

## 3.1 Sequential Pairwise Tests

At one extreme it is obviously simple 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 contributions by Kapetanios (2003, 2008) have provided new techniques that both utilise the power of an omnibus null, with a sequential test that allows greater inference under the alternative hypothesis. However, for a large number of regions locating the partitions over $\digamma$ that are consistent with a particular configuration of convergence clubs generates further difficulties both because the number of combinations is large and related, that we have little prior information.[15] 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 KPSS test. We introduce the test by first denoting $\mathbf{y}_t = \{y_{it}\}$ as the $N \times 1$ vector of log per capita income and write $\mathbf{y}_t$ as

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} t + \mathbf{D} \sum_{s=1}^{t} \mathbf{v}_s + \boldsymbol{\varepsilon}_t, \tag{10}$$

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 $\mathbf{y}_t$, $m \in (0, ..., N)$, and $\mathbf{D} = \{D_{i,l}\} \in \{0, 1\}$ denotes a $N \times m$ matrix. $\boldsymbol{\varepsilon}_t = \{\varepsilon_{(ij),t}\}$ is a $N \times 1$ vector of stochastic components.

[15] 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.

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

$$y_{(ij),t} = \alpha_{(ij)} + \beta_{(ij)}t + D_{(ij),l}\sum_{s=1}^{t} v_s + \varepsilon_{(ij)t}. \tag{11}$$

(11) 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 cointegtaring vector $[1 - 1]$. The additional parameter restrictions for the null hypothesis of absolute convergence are that $\alpha_{(ij)} = 0 \;\; \forall \imath \neq j \in \digamma$. Both asymptotic absolute and relative convergence imply that the cross sectional variance of log per capita income converges to a finite level. Using the notation developed in section 2.3, the test statistics for zero mean and level stationarity are given by

$$\begin{aligned} \widehat{\tau}_0 &= T^{-2}\sum_{t=1}^{T} S_t'[\widehat{\sigma}^2]^{-1}S_t \\ \widehat{\tau}_\mu &= T^{-2}\sum_{t=1}^{T} \bar{S}_t'[\widehat{\sigma}^2]^{-1}\bar{S}_t. \end{aligned} \tag{12}$$

where $S_t = \sum_{t=1}^{t} y_{(ij),s}$. Examining (11), 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 straightforward to test whether two regions form part of a single group. However, for a large number of regions locating the partitions over $\digamma$ 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 $\alpha_{(ij)}$. An alternative testing strategy 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 $\digamma = \{1, 2, 3, 4\}$.

(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, $p_{(1,2)} = \{\max_{i,j\in\digamma}\{\widehat{\mathbf{p}}_{s=1}\} > p_{\mathtt{min}}\}$ then regions 1,2 are the first pair of regions to form a club[16]. 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 $K(1')$.

[16] 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.

(ii) In the second iteration ($s = 2$) we define the set of regions as $F^{'} = (1^{'}, 3, 4)$. 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 $p_{(r,v)} = \{\max_{i,j \in F}\{\widehat{\mathbf{p}}_{s=2}\} > p_{\mathtt{min}}\}$, then if, for example, $p_{(r,v)} = p_{(1'3)}$, the singleton cluster $K(3)$ joins cluster $K(1')$ forming a three-region cluster $K(1^{''}) = \{1, 2, 3\}$.

(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 Kapentanois 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 sections 5 and 6 we utilise this test to investigate the extent of regional convergence within the EU.

# 4 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 (11) 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 $(\varepsilon_{(ij),t})$:

$$y_{(ij),t} = \alpha_{(ij)} + \beta_{(ij)}t + r_t + \varepsilon_{(ij),t} \tag{13}$$

$$r_t = r_{t-1} + v_t, \tag{14}$$

where $r_t = \sum_{s=1}^{t} v_s$ represents the first difference of the stochastic trends 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 $\Psi(L) = 1 + \sum_{s=1}^{\infty} \psi_{(ij),s} L^s$.[17] Under these assumptions $\{\varepsilon_{(ij),t}\}$ has an infinite order autoregressive representation

$$\varepsilon_{(ij),t} = \sum_{s=1}^{\infty} \lambda_{(ij),s} \varepsilon_{(ij),t-s} + u_{(ij),t} \tag{15}$$

where $\Lambda(L) = \Psi(L)^{-1} = 1 + \sum_{s=1}^{\infty} \lambda_{(ij),s} L^s$. Given (13), since $\{\varepsilon_{(ij),t}\}$ is a stationary process, the necessary condition for convergence of regions $i, j$ is that the variance of the random walk error $(\sigma^2_v)$ 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.[18] 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 (13) can be combined to give a reduced form ARIMA(0,1,1) model. In particular, assuming independence of $\varepsilon_{(ij),t}$ and $v_t$ (13) becomes a local component model which, after 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 $\sigma^2_{\eta_{(ij)}} = \sigma^2_{\varepsilon_{(ij)}}/\theta$. This then allow us to use the parametric model for sampling instead of the "unobservable" component model. The parametric nature of the sampling scheme facilitates a zero random-walk variance restriction by imposing a moving average unit root i.e. $\theta = 1$ in the parametric ARMA representation.

The reduced form parameter $\theta$ can be determined 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 component (13) and the ARIMA(0,1,1) model:

$$\theta = \frac{1}{2}\left\{ \frac{\sigma^2_v}{\sigma^2_{\varepsilon_{(ij)}}} + 2 - \left( \frac{\sigma^2_v}{\sigma^2_{\varepsilon_{(ij)}}} + 4\frac{\sigma^2_v}{\sigma^2_{\varepsilon_{(ij)}}} \right)^{1/2} \right\}, \tag{16}$$

where $q = \frac{\sigma^2_v}{\sigma^2_{\varepsilon_{(ij)}}}$ is the signal to noise ratio. Under the stationarity null, namely that regions $i, j$ are converging, the variance of the random walk component $(\sigma^2_v)$ is zero, which in turn implies that $\theta = 1$ in the ARIMA representation.

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

$$\Delta y_{(ij),t} = \sum_{k=1}^{p} \phi_{(ij),k} \Delta y_{(ij),t-k} + \eta_{(ij),t} - \theta \eta_{(ij),t-1}, \tag{17}$$

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

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

where $p$ denotes optimal lag length using the AIC criterion which is determined in order to generate white noise errors $\eta_{(ij),t}$. Based on the equivalence of a reduced form and the component model in second moments, in (17) the AR component represents an approximation to the assumed infinite-order moving average errors to capture the dependence structure in the data, whereas the MA component in (17) follows from the reparametrisation of the structural component model to reproduce the stationarity properties of the data in the ARMA representation. By imposing a moving average unit root in the sampling procedure we can then construct the bootstrap distribution of the test statistic under the null.

The accuracy of the bootstrap test relative to the asymptotic approximations, hinges on the bootstrap sample be drawn independently. Given the presence of a known dependence structure, in this case a stationary ARMA($p$,1) model, we utilise the Recursive Bootstrap.[19] To achieve independent re-sampling from (17) we estimate $\hat{\phi}_{(ij),k}$ and $\hat{\eta}_{(ij),t}$ and generate bootstrap samples of centered residuals $\{\bar{\eta}_{(ij),t}\}_{t=1}^{T}$ where $\bar{\eta}_{(ij),t} = \widehat{\eta}_{(ij),t} - \frac{1}{T-1}\sum_{t=2}^{T}\widehat{\eta}_{(ij),t}$[20].

Given the bootstrapped residuals the $r^{th}$ bootstrap sample for the data $\{\Delta y^{r}_{(ij),t}\}_{t=1}^{T}$ is generated based on the recursive relation[21]

$$\Delta y^{r}_{(ij),t} = \sum_{k=1}^{p} \widehat{\phi}_{(ij),k} \Delta y^{r}_{(ij),t-k} + \bar{\eta}^{r}_{(ij),t} - \bar{\eta}^{r}_{(ij),t-1}. \tag{18}$$

We then recover the level of the series (where in this instance the level denotes the contemporaneous regional difference) directly from (18):

$$y^{r}_{(ij),t} = y^{r}_{(ij),t-1} + \sum_{k=1}^{p} \widehat{\phi}_{(ij),k} \Delta y^{r}_{(ij),t-k} + \eta^{r}_{(ij),t} - \eta^{r}_{(ij),t-1} \tag{19}$$

Defining $h^{r}_{t} = y^{r}_{(ij),t} - \frac{1}{T}\sum_{t=1}^{T} y^{r}_{(ij),t}$, 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

$$\widehat{\tau}^{r}_{(ij),\mu} = T^{-2} \sum_{t=1}^{T} \bar{S}^{r,\prime}_{t} [\widehat{\sigma}^{r,2}]^{-1} \bar{S}^{r}_{t}, \tag{20}$$

where $\bar{S}^{r}_{t} = \sum_{s=1}^{t} h^{r}_{s}$. 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, $\hat{\mathbf{p}}^{B}$.

[19]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.

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

[21]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.

# 5 Convergence in the EU

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 and interpretation 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. To inform policy more information is needed on the economic forces that drive the formation of regional convergence clusters. To provide such information we first locate the convergence groups and then confront the resulting cluster compositions, for both the asymptotic and bootstrap tests, with a set of hypothetical clusters generated by hypotheses constructed using a set of economic, socio-demographic, and political indicators suggested by the new economic geography type of models.[22] By looking at the correlation between the cluster patterns suggested by theory and the observed regional convergence clusters we are able to detect the dominant economic forces that explain the formation of the clusters and, related, we can assess whether there is a significant difference between the correlations coefficients in the two versions of the test.

## 5.1 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 note that we are aware of the problems that surround the choice of which spatial units to use. 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. For example, many of the

[22]For example, one hypothesis is that regional convergence takes the form of a core-periphery dichotomy, with regions in the core converging to a different long-run steady state per capita output from those in the periphery.

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. The principal reason for their comments is that 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 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[23] used in this study is given in Table 1.

We use regional data on Gross Value Added[24] 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 *et al.* (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. Table 2 describes the features of the indicators used in the formation of the generated cluster patterns informed by economic theory, which are used to interpret the cluster outcomes (see section 6.2). We classify indicators according to whether they represent *geographical*, *socio-demographic* or *political* factors. All indicators are central components of the new economic geography growth models since they justify the presence of increasing returns and comparative advantage at the sectoral and/or regional level (Fujita *et al.,* 1999; Fujita and Thisse, 2003).

The first set of *geographical* factors classifies regions on the basis of country-membership, a periphery-core distribution of the regions, geographic location and the intensity of the transportation network. In their earlier work on regional convergence, Barro and Sala-i-Martin (1997) argued that regional convergence is more likely amongst regions within a given nation than it is between regions in different nations. Their argument is that institutional frameworks, regulatory systems, consumer tastes, and technologies are much more similar across regions within a given country than they are between different countries. This line of reasoning would lead us to hypothesize a significant country (national) effect on regional convergence clustering. At the same time, recent work on the application of endogenous growth theory to regional development suggests that growth effects arising from knowledge creation and spillovers, on the one hand, and the

[23]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.

[24]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.

accumulation of skilled human capital on the other, tend to exhibit spatial concentration. Strong spatial proximity effects are held to operate, implying a significant degree of spatial dependence in the geographical pattern of growth performance. In other words, we should expect convergence clusters to comprise sets of neighboring or spatially proximate regions.

Another important factor for the location of activity is the intensity of the transportation network. Since production in our four sectors differs in the intensity of transportation costs and in their relative distance from final markets, then regions with a better transport infrastructure might be expected to attract sectors which produce transport intensive commodities. This approach is developed in a trade theory framework in Louveaux *et al.* (1982) and Fujita and Thisse (2002).

On a larger geographical scale, it is often argued that the regional patterns of growth and development in the EU are characterized by a strong and persistent core-periphery structure, in which a core of leading growth regions encompassing the South East region of the UK, parts of the Netherlands, the Paris region, the Brussels region, Southern Germany, and Northern Italy, is contrasted with a periphery of slower growing regions. The implication is that regional convergence dynamics should reflect a core-periphery dichotomy.

The second set of *socio-demographic* factors classifies regions on the basis of population growth and agglomeration effects. Along these lines Martin and Ottaviano (2001) show that growth and geographical agglomeration are self-reinforcing processes. In fact, agglomeration increases with growth since it is always more convenient to locate the activity where the final market is bigger or the production of knowledge is higher. At the same time growth increases with agglomeration since agglomeration reduces the cost of innovating in those regions where economic activity concentrates.

Finally, the third set of *political* factors classifies regions on the basis of political intervention (within the EU) which are designed to encourage and guide structural adjustment of poorer regions. The instruments used include the European Development Fund, the European Social Fund and the European Agricultural Guidance and Guarantee Fund (Martin and Tyler, 2000).

# 6 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 3 summarises this information in terms of the number and size of the convergence clubs and group characteristics, such as average per-capita income. To facilitate the interpretation of our results we compare the convergence outcomes against a set of cluster patterns generated by economic indicators suggested by the New Economic Geography.

## 6.1 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 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 3 we present the frequency distribution of the cluster size for both bootstrap and asymptotic tests and for each[25] 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 and lower 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 pattern, $\mathbf{M}^B$, is then given by

$$\zeta_l = \left( \frac{\sum_{i=1}^{N} \sum_{j \neq i}^{N} m_{ij}^B \times m_{ij}^A}{\left( \sum_{i=1}^{N} \sum_{j \neq i}^{N} m_{ij}^B \right)^{1/2} \left( \sum_{i=1}^{N} \sum_{j \neq i}^{N} m_{ij}^A \right)^{1/2}} \right)^{1/2}, \tag{21}$$

where $l$ indexes the set {Agriculture, Manufacturing, Market Services, Non-Market Services}. The results are reported in panel III of Table 3. With correlation coefficients ranging between 50%

[25] 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.

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.

**Mean Income** In order to assess the properties of each cluster we compute mean log per-capita income, [26] $\bar{x}_g$ for each test. A visual impression of the oversized property of the asymptotic test of convergence is evident in the relative kurtosis of this distribution, presented in Figure 4. This results from an overrejection of the convergent null, thereby generating a distribution with a large number of small clubs with higher average income. Examining the comparable bootstrap distribution, shows a marked decrease in kurtosis and a commensurate narrowing of the gap between the richest and the poorest cluster[27]. Summary statistics are provided in the last three columns of panels I and II of Table 3. Note that for the bootstrap distribution some of the richest clusters of smaller size are absorbed into clusters with lower mean per-capita income which is manifest 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-per capita log income of the *poorest* cluster, $\bar{x}_{min}$, by around 24% (from 9.4 to 11.7) and a decrease in mean per-capita income of the *richest* cluster, $\bar{x}_{max}$, 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 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.

## 6.2 Interpreting Convergence Clusters

To date we have demonstrated that tests of convergence based on bootstrap and asymptotic methodologies generate substantially different outcomes in terms of the distribution of convergence clubs. 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. However, as a consequence it is often difficult to interpret the results, and in particular understand the forces which are consistent with the observed convergence clubs which is important to inform policy decisions. Below we assess the extent to which the generated cluster patterns are consistent with one or more of the artificially constructed cluster patterns suggested by the new economic geography theory.

We collect the hypothetical $(h)$ clusters in a $N \times N$ matrix $\mathbf{M}^h = \{m_{ij}^h\}$, with typical element $m_{ij}^h$ equal to 1 if regions $i$ and $j$ belong to the same cluster and zero otherwise. For example, one hypothesis is that regional convergence takes the form of a core-periphery dichotomy, with regions in the core converging to a different long-run steady state per capita output from those in the periphery. Based upon this hypothesis, it is possible to construct a matrix with cell entries $m_{ij}$ either zero or one; $m_{ij} = 1$ indicates that regions $i$ and $j$ are part of a core region.

The correlation parameter $(\zeta_l^h)$ between the constructed, $\mathbf{M}^h$, and the observed cluster pattern,

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

[27] Comment on different scales and non-overlapping distributions, and why different scales

$\widehat{\mathbf{M}}$, is given by:

$$\zeta_l^h = \left( \frac{\sum_{i=1}^{N} \sum_{j \neq i}^{N} m_{ij}^h \times \widehat{m}_{ij}}{\left( \sum_{i=1}^{N} \sum_{j \neq i}^{N} m_{ij}^h \right)^{1/2} \left( \sum_{i=1}^{N} \sum_{j \neq i}^{N} \widehat{m}_{ij} \right)^{1/2}} \right)^{1/2} \qquad (l = B,\ A), \tag{22}$$

where $A$ and $B$ index, respectively, asymptotic and bootstrap outcomes. In testing whether the correlation between the observed and generated clusters are different for the two tests we first convert each correlation coefficient into a $Z$-score using Fisher's $Z$ transformation. [28] We then use the statistic $z^h = \frac{Z_B^h - Z_A^h}{\sigma_{z_B^h - z_A^h}}, \quad l = A, B.$ to test whether there is a significant difference in the two correlation coefficients.[29]

Below we consider the correlation of the cluster outcomes with the patterns generated by each of the following factors: (i) *geography* , (ii) *socio-demography*, and (iii) *political* (see Table 2). Table 4 reports the correlation results for agriculture, manufacturing, market and non-market services. A positive (negative) value of the Z-score statistic indicates that the correlation coefficient between the generated and the hypothesised cluster patterns is greater (lower) for the bootstrap test. For all three types of clustering hypotheses – geographical, socio-demographic and political – we observe a positive Z-score statistic. In terms of geography, with the exception of manufacturing, the bootstrap convergence clubs exhibit closer correspondence with hypothesised clusters derived on the basis of country membership, reflecting the importance of national level effects on regional growth patterns. For the agriculture and non-market services sectors there is evidence of a closer correspondence between observed and hypothesised clusters based on a core-periphery distribution of regions. For all sectors other than non-market services, a cluster pattern based on geographical contiguity is also more correlated with the cluster distribution generated by the bootstrap test. For agriculture we observe a statistically significant increase in the correlation between our observed clusters and groupings defined on the basis of local relative specialisation in agriculture.

Socio-demographic factors, specifically population density and settlement structure, are also important in interpreting the results. The cluster outcomes for non-market services correlate with the clusters based on population density. This correlation is significantly higher when we consider the bootstrap version of the test, suggesting that local market-demand factors and dense labour markets may be of importance for this sector. Turning to the political factors, the bootstrap results confirm the asymptotic results and we find little evidence that regional convergence has been influenced by the provision of the EU Structural and Cohesion Funds.[30] In grouping regions according to their Objective funding status, only the agricultural sector exhibits a significantly

[28]The Fisher's transformation is defined as $Z_l^h = \frac{1}{2} \ln \frac{1+\zeta_l^h}{1-\zeta_l^h}$.

[29]The statistic $z$ is normally distributed with standard deviation $\sigma_{Z_B - Z_A} = \sqrt{\sigma^2_{Z_B^h} + \sigma^2_{Z_A^h}}$. Since the sample size for the two periods are equal then $\sigma_{z_B^h - z_A^h}$ is equal to $\sqrt{\frac{2}{N(N-1)-3}}$ (see Cohen and Cohen, 1983).

[30]Our result is in line with other empirical evidence on the impact of structural funds on targeted European regions (see Dall'Erba and Le Gallo (2008) which shows that convergence takes place but that funds have no impact on it.

positive Z-score statistic. Other studies have found mixed evidence that EU regional policy has contributed to regional convergence (see, for example, Braunerhjelm *et al.,* 2000; Boldrin and Canova, 2001; Puga, 2002), and our results tend to confirm this ambiguity.

## 7 Conclusions

With the increasing dimension of both country and regional datasets, economists have, in principle, greater opportunity to consider the question of convergence at a finer scale. In this context we have reviewed recent trends in testing strategies and observed a definite trend away from the application of large omnibus null tests of hypotheses with the commensurate problems of dimensionality and very limited inference under the alternative,

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. 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 across regions within the European Union for a number of industrial sectors.

To further assess the driving forces behind the convergence clusters across the four sectors, our observed clusters were then compared with a number of hypothesized regional groupings based on different theories and models of regional growth and development. We provide estimates of the correlations between our observed outcomes and these cluster patterns. For all three types of clustering hypotheses – on the basis of location and socio-demographic factors, and policy status – there is a tendency for the correlation between the cluster types suggested by new economic geography and our observed clusters to increase from the asymptotic to the bootstrap test. Hence, by resolving the size distortion which afflicts the asymptotic test we are not only able to find more convergence but also to find stronger support for the economic forces that drive the formation of the regional convergence clusters.

## References


[1] Abramovitz, M. (1986). 'Catching Up, Forging ahead, and Falling behind', *Journal of Economic History*, 46(2), pp. 385-406.

[2] Barro, R. J. and Sala-i-Martin, X. (1991). 'Convergence across States and Regions', *Brookings Papers on Economic Activity*, 1, pp. 107-82.

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

[4] Benjamini, Y. and Hochberg, Y. (1995). 'Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society Series B*, 57, pp. 289-300.

[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] Bertola, G. (1993). 'Models of Economic Integration and Localized Growth', in (F. Torres and F. Giavazzi, eds.), *Adjustment and Growth in the European Monetary Union*, pp. 156-79, Cambridge: Cambridge University Press.

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

[9] 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.

[10] Braunerhjelm, P., Faini, R. Norman, V., Ruane, F. and Seabright, P. (2000). 'Integration and the Regions of Europe: How the Right Policies Can Prevent Polarization'. London: Centre for Economic Policy Research.

[11] 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.

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

[13] Canova, F. (2001). 'Are EU Policies Fostering Growth and Reducing Regional Inequalities?', CREI Working Paper No. 8, University Pompeu Fabra.

[14] Canova, F. (2004). 'Testing for Convergence Clubs in Income per Capita: a Predictive Density Approach', *International Economic Review*, 45(1), pp. 49-77.

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

[16] Cheshire, P. C. and Hay, D. G. (1989). *Urban Problems in Western Europe: An Economic Analysis*. Unwin Hyman, London.

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

[18] Cohen, J. and Cohen, P. (1983). *Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Hillsdale, NJ: Erlbaum.

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

[20] Dall'Erba, S. and Le Gallo, J. (2008). 'Regional Convergence and the Impact of European Structural Funds over 1989–1999: A Spatial Econometric Analysis', *Papers in Regional Science*, 87(2), pp 219-244.

[21] Deckers, T. and C. Hanck (2009). 'Multiple Testing Tecniques in Growth Econometrics', Munich Personal RePEc Archive Paper No. 17843.

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

[23] 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*, Volume 1A, North-Holland: Amsterdam, pp. 555-677.

[24] 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.

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

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

[27] 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.

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

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

[30] Fujita, M. and Thisse, J. F. (2003). 'Agglomeration and Market Interaction', in (M. Dewatripont, L.P. Hansen and S.T. Turnovsky, eds.), *Advances in Economics and Econometrics. Theory and Applications. Eight World Congress*, vol.1, pp. 302-38, Cambridge: Cambridge University Press.

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

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

[33] 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.

[34] Hjort, N.L. and Claeskens, G. (2003). 'Frequentist Model Average Estimators', *Journal of the American Statistical Association*, 98, pp. 879-899.

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

[36] Hockhberg, Y. (1988). 'A Sharper Bonferroni Procedure for Multiple Tests of Significance', *Biometrika*, 75, 800-802.

[37] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). 'Bayesian Model Averaging: A Tutorial (with Discussion)', *Statistical Science*, 14, pp. 382–401.

[38] Holmes, M. J. (2000). 'Convergence in International Output: Evidence from Panel Data Unit Root Tests', Business Cycle Volatility and Economic Growth Research Report No. 00/6, Loughborough University.

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

[40] 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.

[41] Johansen, S. (1988). 'Statistical analysis of cointegration vectors', *Journal of Economic Dynamics and Control*, 12, pp. 231-54.

[42] Jones, H. E., Ohlssen, D. I. and Spiegelhalter, D. J. (2008). 'Use of the False Discovery Rate when Comparing Multiple Health Care Providers', *Journal of Clinical Epidemiology*, 61, pp. 232-240.

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

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

[45] 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.

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

[47] 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.

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

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

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

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

[52] 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.

[53] Martin, P. and Ottaviano, G. I. P. (2001). 'Growth and Agglomeration', *International Economic Review*, 42(4), pp. 947-68.

[54] Martin, R. and Tyler, P. (2000). 'Regional Employment Evolutions in the European Union: a Preliminary Analysis', *Regional Studies*, 34(7), 601-16.

[55] 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.

[56] Moon, H. R. and Perron, B. (2009). 'Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Non Stationarity Properties of Individual Series in a Panel', Centre Interuniversitaire de Recherche en Economie Quantitative, CIREQ Cahier 10-2010.

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

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

[59] Phillips, P. C. B. and S. Ouliaris (1988). 'Testing for cointegration using principal components methods', *Journal of Economic Dynamics and Control*, 12, pp. 205-230.

[60] Puga, D. (2002). 'European Regional Policy in Light of Recent Location Theories', *Journal of Economic Geography*, 2(4), pp. 373-406

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

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

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

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

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

[66] 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.

[67] 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: 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 2: Geographic, Socio-demographic, and Political Indicators

| | Factors | Description |
|---|---|---|
| Geographical | | |
| | Country membership | Regions cluster solely on the basis of their nation-state membership. The associated mechanisms include a shared institutional framework and a well defined geographic boundary. |
| | Core-Periphery | Regions are classified according to their relative distance with respect to a core of European regions. |
| | Geographic location | Regional clusters are determined by a broader geographical classification of regions: Northern European, Atlantic, Mediterranean, Central or Eastern European. Here, it is assumed that contiguity and institutional similarity may affect regional convergence. |
| | Transportation network by total area | Regions are classified according to the intensity of the transportation network. |
| | Agricultural intensification | Regions are classified according to a composite indicator of percent growth of agricultural accounts, percent of agricultural holdings greater than 50% and percentage of land use by total area. |
| Socio-demographic | | |
| | Population growth by area | Regions are classified according to the average of population growth between 1991 and 1995. Changes in population growth and population density capture the role of urban agglomeration in shaping real GVA per capita convergence. |
| | Settlement structure | Regions are classified according to the number of inhabitants and population density. This may reflect, for example, different levels of urbanization and agglomeration dynamics. |
| Political | | |
| | EU Structural Funds Objectives | Regions are classified according to the following EU Cohesion and Structural Fund objectives:<br>Objective 1. To promote the development and structural adjustment of underdeveloped regions.<br>Objective 2. To redevelop regions or areas within regions seriously affected by industrial decline.<br>Objective 3. To combat long term unemployment, to provide career prospects for young people (aged under 35) and to reintegrate persons at risk of being excluded from the labour market.<br>Objective 4. To facilitate the adoption of workers to industrial change and developments in the production system.<br>Objective 5a. To speed up the adoption of production, processing and marketing structures in agriculture and forestry and to help modernize the fisheries and aquaculture sector.<br>Objective 5b To promote the development of rural areas.<br>Objective 6. To promote the development of northern regions in the new member states in Scandinavia -since 1995 Finland and Sweden. |

Table 3: Joint Frequency Distribution

I: Asymptotic

| Cluster size | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Total Clusters | $\sigma_{\bar{x}}$ | $\bar{x}^{min}$ | $\bar{x}^{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of Clusters | | | | | | | | | | | Summary Statistics | | |
| 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

| Cluster size | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | $\sigma_{\bar{x}}$ | $\bar{x}^{min}$ | $\bar{x}^{max}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of Clusters | | | | | | | | | | | | | |
| 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

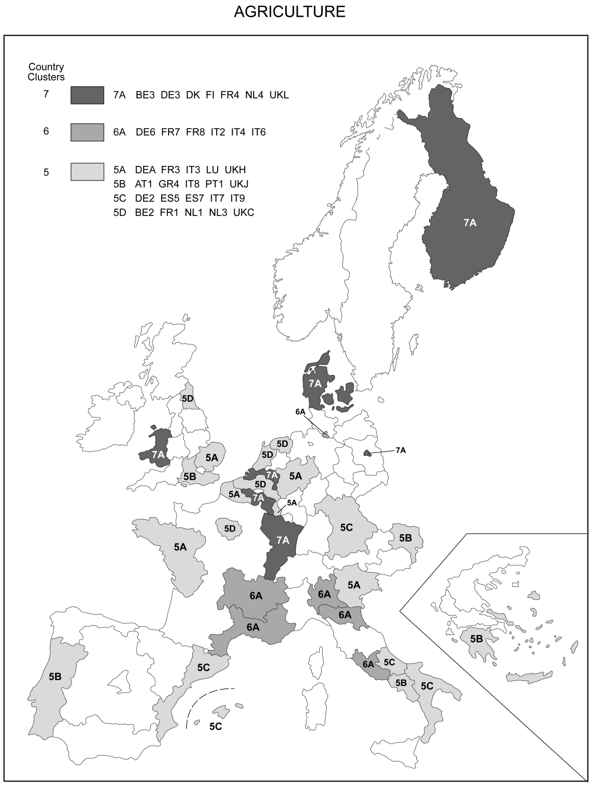
| | |
|---|---|
| Agriculture | 0.672 |
| Manufacturing | 0.509 |
| Market Services | 0.557 |
| Non-Market Services | 0.591 |

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

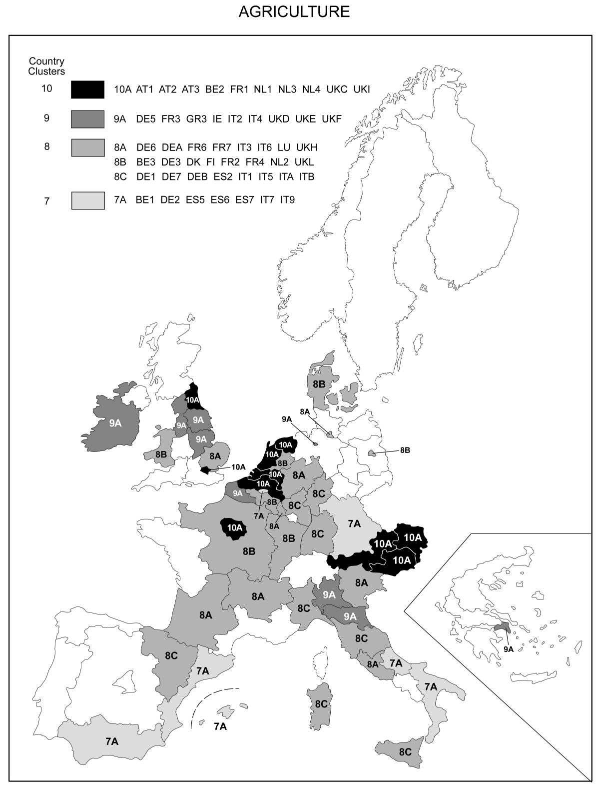
Table 4: Correlation Between Observed and Cluster Patterns Informed by NEG Theory: Univariate Analysis

| | | Agriculture | Manufacturing | Market Services | Non-Market Services |
|---|---|---|---|---|---|
| *Geographical* | | | | | |
| | Country Membership | | | | |
| | Asymptotic | 0.299 | 0.345 | 0.345 | 0.398 |
| | Bootstrap | 0.388 | 0.317 | 0.465 | 0.493 |
| | $z$ | $(4.89)^{**}$ | $(-1.52)$ | $(6.97)^{**}$ | $(5.75)^{**}$ |
| | Core-Periphery | | | | |
| | Asymptotic | 0.354 | 0.295 | 0.271 | 0.364 |
| | Bootstrap | 0.417 | 0.332 | 0.341 | 0.340 |
| | $z$ | $(3.59)^{**}$ | $(1.99)^{*}$ | $(3.74)^{**}$ | $(-1.33)$ |
| | Geographic Location | | | | |
| | Asymptotic | 0.321 | 0.273 | 0.364 | 0.356 |
| | Bootstrap | 0.381 | 0.322 | 0.398 | 0.380 |
| | $z$ | $(3.32)^{**}$ | $(2.60)^{**}$ | $(1.93)^{*}$ | $(1.34)$ |
| | Transportation Network | | | | |
| | Asymptotic | 0.330 | 0.341 | 0.272 | 0.343 |
| | Bootstrap | 0.369 | 0.387 | 0.321 | 0.356 |
| | $z$ | $(2.15)^{**}$ | $(2.57)^{**}$ | $(2.60)^{**}$ | $(0.72)$ |
| | Agricultural Intensification$^{\dagger}$ | | | | |
| | Asymptotic | 0.373 | | | |
| | Bootstrap | 0.437 | | | |
| | $z$ | $(3.71)^{**}$ | | | |
| *Socio-Demographic* | | | | | |
| | Population Growth by Area | | | | |
| | Asymptotic | 0.314 | 0.315 | 0.297 | 0.369 |
| | Bootstrap | 0.363 | 0.301 | 0.373 | 0.404 |
| | $z$ | $(2.68)^{**}$ | $(-0.75)$ | $(4.15)^{**}$ | $(1.99)^{*}$ |
| | Settlement Structure | | | | |
| | Asymptotic | 0.365 | 0.338 | 0.363 | 0.390 |
| | Bootstrap | 0.433 | 0.403 | 0.416 | 0.450 |
| | $z$ | $(3.92)^{**}$ | $(3.65)^{**}$ | $(3.03)^{**}$ | $(3.53)^{**}$ |
| *Political* | | | | | |
| | EU Structural Fund Objectives | | | | |
| | Asymptotic | 0.367 | 0.296 | 0.389 | 0.367 |
| | Bootstrap | 0.428 | 0.326 | 0.373 | 0.376 |
| | $z$ | $(3.45)^{**}$ | $(1.58)$ | $(-0.89)$ | $(0.49)$ |

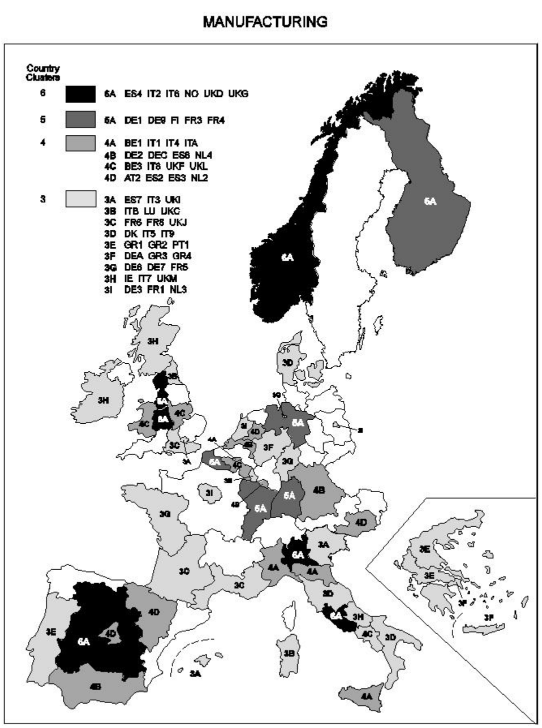
** (*) denotes significance at the 5% (10%) level. $^{\dagger}$ Data are available only for the agricultural sector.
Note: All correlation coefficients for each period were tested and were found to be significantly different from zero at the 5% level.



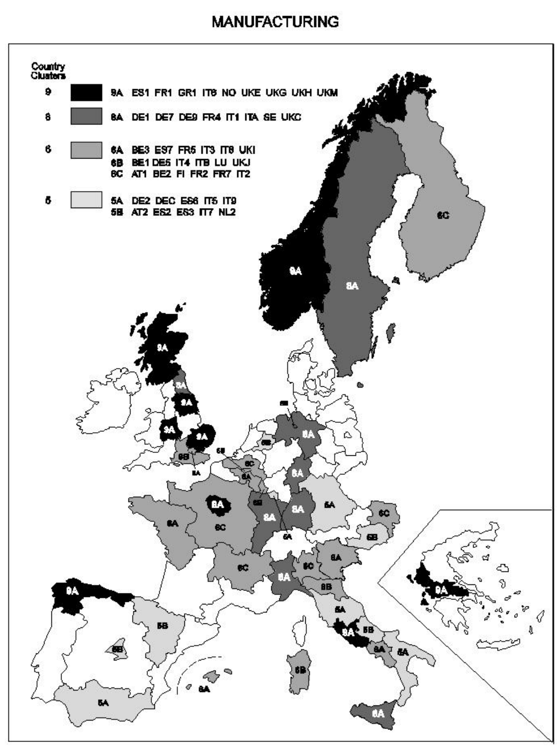

(a) Relative Convergence in Manufacturing: Asymptotic Results




(b) Relative Convergence in Manufacturing: Bootstrap Results

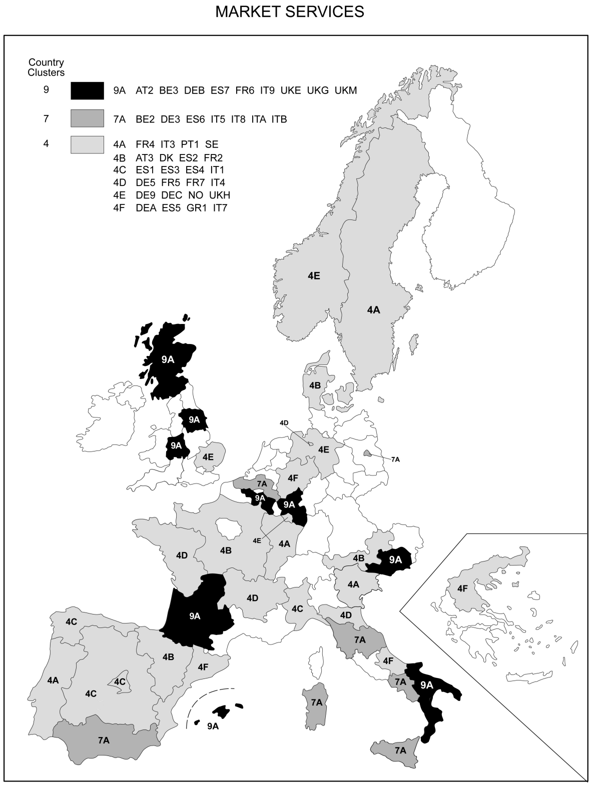



(c) Relative Convergence in Manufacturing: Asymptotic Results

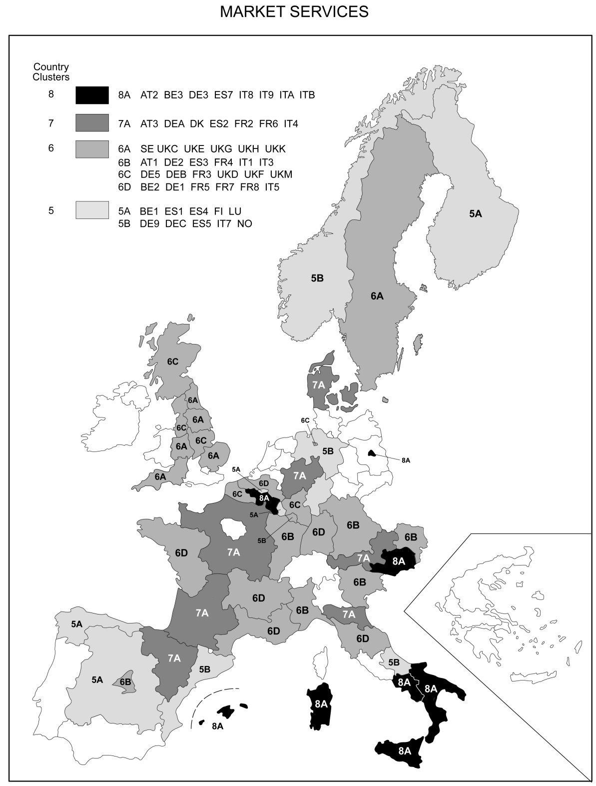



(d) Relative Convergence in Manufacturing: Bootstrap Results

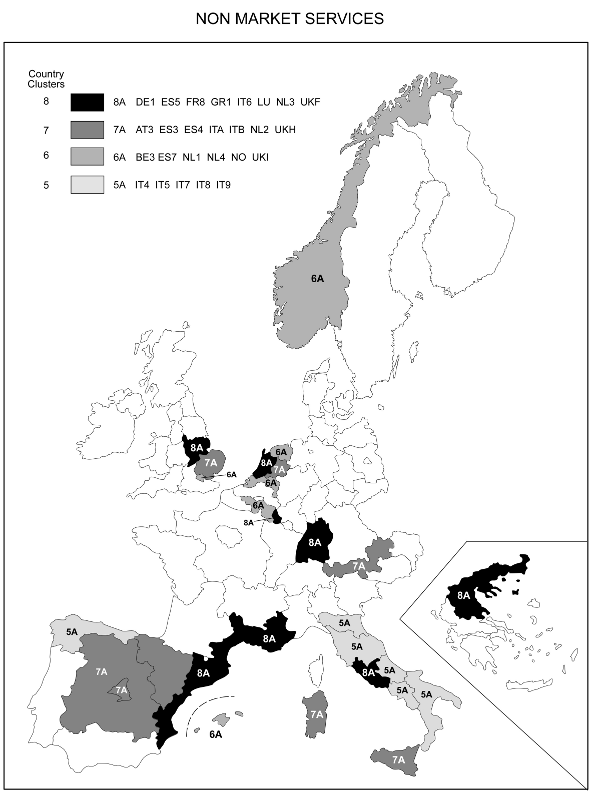
Figure 1: Asymptotic and Bootstrap Results for Agriculture and Manufacturing



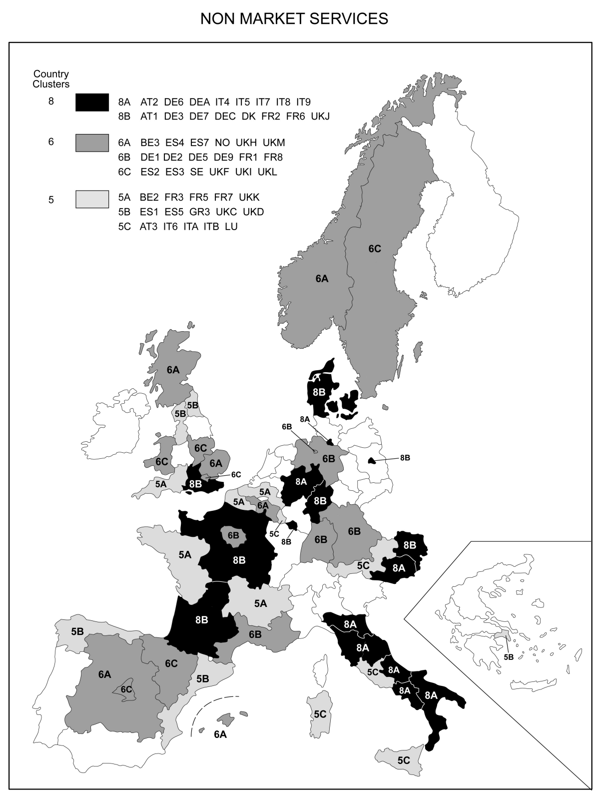

(a) Relative Convergence in Market Services: Asymptotic Results




(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

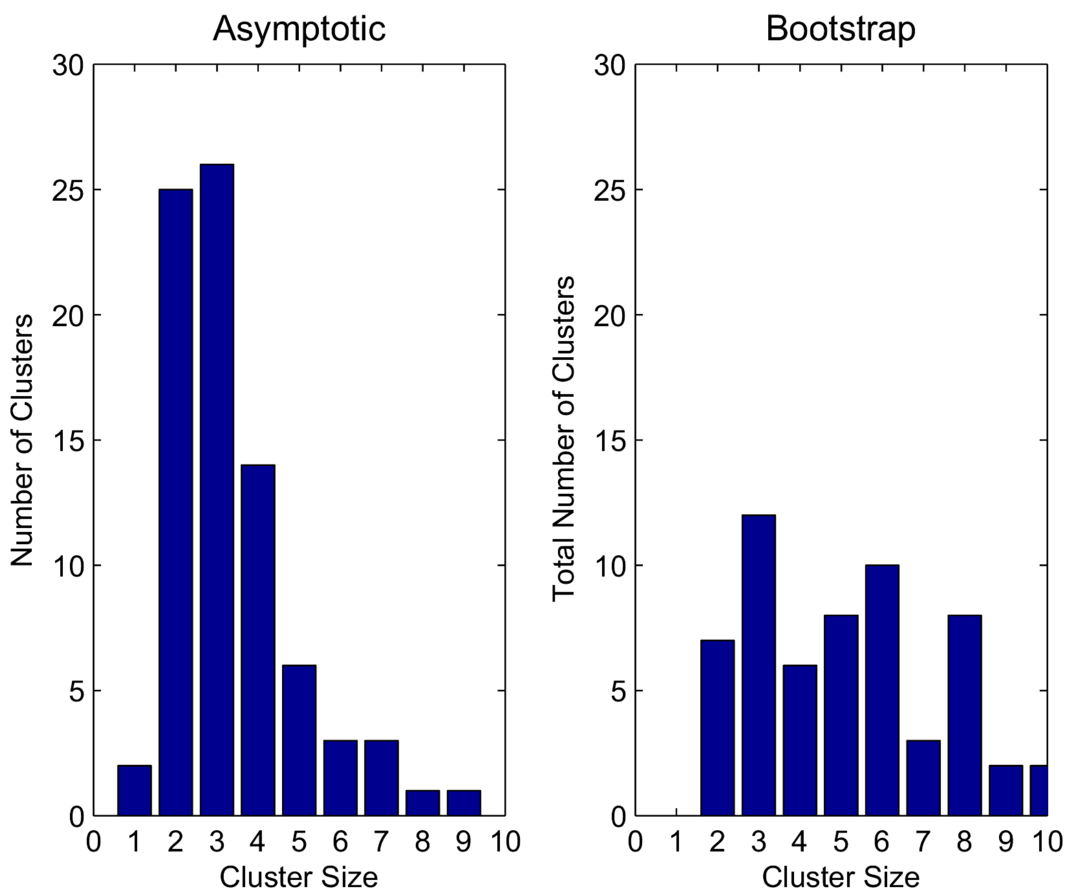



Figure 3: The Distribution of Cluster Size.

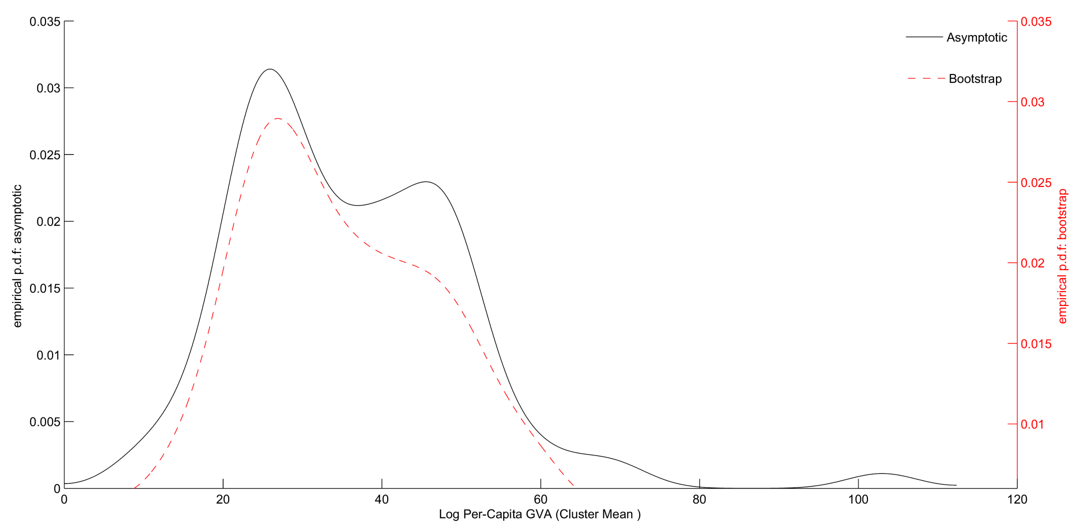



Figure 4: The distribution of average log per-capita GVA by cluster: All sectors.

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