Empirical Evidence of Income Dynamics Across EU Regions

Maria Grazia Pittau and Roberto Zelli

University of Rome "La Sapienza"

Abstract

This paper analyses the distribution of purchasing power standardised per capita income across EU-12 regions between 1977 to 1996. Dispersion of incomes between regions is measured taking into account their population sizes. The cross-sectional distributions are initially described by weighted kernel density estimates, revealing a multimodal structure of the distributions, less evident over the period. This evidence is supported by a bootstrap test. To detect homogeneous groups of regions, the empirical distributions are approximated by finite mixture of normal densities. The components of the mixture represent clusters of poor/rich regions, while the mixing proportions the allocation over the poor and the rich components. The number of components is assessed by a bootstrap LR test, while the goodness of fit by a kernel-density-based test. Income mobility is modelled by the stochastic kernel, the continuous counterpart of the transition probability matrix. The main implication is a very slow process of catching up of the poorest regions with the richer ones and a process of shifting away of a small group of very rich regions. This evidence is reflected in the shape of the ergodic distribution which is well fitted by a two-component mixture model.

Key words: *kernel estimation; mixture models; stochastic kernel; regional convergence; EU income disparities.*

JEL classification: C14, O11, O52.

Acknowledgements. The authors gratefully acknowledge financial support from University "La Sapienza", research grant n. C26F024479. We would like to thank the co-editor Steven Durlauf, three anonymous referees and participants of the XXVII IARIW conference, Djurhamn, Sweden, August 2002, for their useful comments and suggestions. Obviously we are solely responsible of any further error and omission.

Address for correspondence: Roberto Zelli, University of Rome "La Sapienza" DCNAPS, Box 83 Roma 62, P.le Aldo Moro 5, 00185 Roma, Italy. Email: roberto.zelli@uniroma1.it; Tel. +390649910834; Fax. +39064453246

1 Introduction

The equalisation of per capita income levels in the European Union is a target that has been made explicit in the Treaty on European Union, documented by Article 2 and, more particularly, by Article 158 (ex 130a). An increasing volume of the EU budget has been devolved towards this objective. From 1975 to 1988 this has been carried out through the European Regional Development and European Structural Funds; subsequently implemented through the European Social Fund (ESF) from 1989, and additionally from 1994 through the Cohesion Fund. Much effort has been devoted to measuring and explaining economic disparities among European regions over time, applying various statistical tools. The debate ramped up dramatically as a result of some empirical evidence showing a tendency of stagnation of regional disparities, during the 1980s and 1990s, casting some doubts on the effectiveness of the EU policy.

This controversial issue comes up for debate in the economic literature on convergence across economies. Cross-section regressions have been the traditional tools of the empirical analysis, which is based on the hypothesis of convergence towards a steady-state growth path. Running the so called growth initial level regressions, possibly including additional variables, a negative coefficient (β) associated to the initial income variable peaks up the negative correlation between the initial income level and the growth rate, consistent with a scenario, where economies that start out poor tend to grow faster, on average, than economies that start out rich[1]. The β approach has been widely utilised to give information on structural parameters of neoclassical type growth models. However, this approach has been criticised essentially since, as it focuses only on an average behaviour, it is uninformative in relation to distribution dynamics. According to this criticism, it is more relevant to capture the dispersion of the cross-sectional distribution of income over time. Such dispersion is typically measured by the standard deviation of the distribution of log incomes across economies, the σ-convergence[2]. While intuitively simple and attractive, σ-convergence does not provide insights on the relative movements of individual economies within the income distribution. Moreover, interpreting measures of dispersion is not straightforward when distributions are not unimodal.

Consequently, a *model of explicit distribution dynamics* has been advocated by several

[1] Also panel data studies on convergence and time series approach deal with the analysis of β convergence. For a comparison and a criticism of the cross section and time series approaches to test convergence hypothesis, see Bernard and Durlauf (1996).

[2] This is the oldest notion of convergence that draws back to the work of Kuznets (1955).

authors, (see, in particular, Quah 1993; 1996a; 1997). Distribution dynamics can be evaluated according to twofold aspects: shape dynamics and intra-distribution mobility. Aim of the paper is to explore how the entire distribution of regional per capita income evolves over time in the European Union, addressing both these features. The first concerns the degree to which the location and shape of the distribution change over time. A natural approach to assess these shape dynamics is to estimate the cross-sectional distributions by using non parametric smoothing procedures, like kernel estimates, which avoid the strong restrictions imposed by traditional parametric estimation. In this framework if there is a bimodal density at a given point of time, indicating the presence of two groups in a population of regions (say a group of "poor" and a group of "rich"), convergence implies a tendency of the distribution to progressively move towards unimodality. The analysis based on kernel density, however, relies heavily on visual impression. Specific aspects of the shapes, that otherwise would be undetected, can be highlighted by non parametric statistical tests. In this paper, the presence of a multimodality structure in the fitted density is assessed by a bootstrap test as in Bianchi (1997). Moreover, to discriminate between significant and not significant shifts in the distributions, a test of closeness between unknown distributions functions is performed.

Another approach that reveals some aspects of the distribution dynamics is to model the per capita income by a finite mixture density. The main advantages of mixture modelling are, the greater flexibility and precision in modelling the underlying distributions of income data, especially when the non parametric density exhibits more than one mode, as well as the possibility to relate the estimated distributions directly to subpopulations which reflect economically homogeneous groups of regions. In this context, the process of convergence occurs when the distributions are better approximated over time by a smaller number of components[3]. Since for the European income distribution there is no *a priori* about the number of subpopulations, the assessment of the smallest number of components compatible with the data is one important issue addressed in this paper. The choice of the components is found out by a bootstrap LR test. How well the estimated mixtures fit the data is further established by a goodness of fit test based on kernel method, extending the results of Fan (1994).

Shape dynamics does not directly address the mobility pattern within the income distribution, i.e. which units move up and down the income ladder. Following Quah (1993), the evolution of the distribution of incomes across regions can be modelled using a stationary first-order Markov chain. This implies the estimation of a probability matrix of transition between quantiles and, in general, discrete states. The results in such cases

[3]If the number of components does not change over time and under the restriction of constant proportions, convergence takes place when the distance between the component means reduce and the component variances does not increase (see Paap and van Dijk, 1998; Tsionas, 2000)

are known to be sensitive to the arbitrary grouping of observations used to discretise the data (Reichlin, 1999; Bulli, 2001). The stochastic kernel, the continuous counterpart of the transition probability matrix, is indicated as a suitable tool to overcome this problem (Quah, 1996c, 1997, 2001). In this analysis, income dynamics over τ periods of time from a given income value in period t, is modelled by the stochastic kernel. Under the assumption of stationarity of the underlying process, the long run limit of the distribution of incomes across regions is estimated. This ergodic distribution is also interpreted in terms of mixture of components.

The evidence of European regional income distribution usually concerns the behaviour of incomes in terms of regions. While this is a common way to view the data, it can be misleading when the focus is on equity and on income gaps between European citizens. A more appropriate approach is to weigh per capita income of each region proportionally to its population, with the assumption of equal distribution within each region. The unit of observation, therefore, becomes a person and not a region. Following the suggestion of a referee, population-based weights are integrated in all the estimation procedures.

The rest of the paper proceeds as follows. The next section reviews some conclusions from theories on economic growth and regional development. The selection of the statistical units and the data-set is discussed in Section 3. The empirical results of the kernel density estimates of the per capita (log) income are reported in Section 4, along with the multimodality and the changing shape tests. Section 5 introduces mixture models, gives details on the choice of the number of components and describes the empirical fitting and the evolution of the probabilistic clusters of regions generated by the components of the mixtures. An analysis of the goodness of fit concludes this section. Section 6 concisely explains the statistical methodology of the stochastic kernel, and discusses the main empirical results. Finally, some concluding remarks are given in Section 7.

2 Interpreting convergence hypothesis

The basic neoclassical growth model predicts that an economy will move toward a unique, globally stable, equilibrium, regardless of its initial conditions, due to diminishing returns to capital. In terms of cross-regional comparison, this means absolute convergence, that is differences in per capita incomes of regions are only transitory. The only difference across regions lies in their initial endowment of capital and poor economies grow faster than the rich ones simply because they are more distant from their steady state.

If each economy approaches its own but unique steady-state equilibrium, then neoclassical theory predicts conditional convergence. Levels of per capita incomes of those economies with similar structural characteristics converge in the long run, independently from their initial conditions. Therefore, poor economies grow faster only if differences in these fundamentals are controlled for. Accordingly, policy interventions to correct initial income differentials are viewed as unnecessary.

In contrast, the idea of club convergence is based on models that yield multiple equilibria. Economies converge to each other if also their initial conditions are in the basin of attraction of the same steady state. For example, the model proposed by Azariadis and Drazen (1990) shows that the asymptotical income distribution tends to polarize because of the presence of externalities that yield social increasing returns to scale only once a threshold level of human capital is reached. The initial conditions, therefore, determine whether an initially poor country will remain stuck in a development trap or will join the club of rich nations (Desgoits, 1999)[4].

Convergence across economies has also been debated in endogenous growth theories. The main features of knowledge (non rivalry, partial excludability, accumulability) imply unbounded accumulation and spillovers effects. Production of knowledge, even if originating in a specific country or region, gradually moves beyond its boundaries, contributing to boost productivity growth in other areas. Convergence may occur when spillovers flow freely between regions and countries. If not, persistent and even widening levels of regional income disparities are allowed. The presence of research and developments (R&D) spillovers may give place to a particular form of club convergence, in which only a selected group of economies performing R&D tends to reach the same growth rate in the long run, while all the others stagnate, as they cannot take advantage of technology transfer (Howitt, 2000). A key point is the ability of backward regions to adapt to and to imitate knowledge developed in the advanced regions. Diffusion stimulates a process of catching up, while constant innovation of the leaders increases the gap to fill for followers (Maurseth, 2001). Recent theories of economic geography, which show that geographical patterns interact with growth processes, tend to explain how physical distance influences regional technology spillovers (see, e.g., Martin and Ottaviano, 2001; Baldwin *et al.*, 2001, Caniels and Verspagen, 2001).

Two remarks are worth emphasising. First, in general, theories on convergence deal with output per worker and the conclusions reached for labour productivity can be maintained also for per capita income only adding the hypothesis that both unemployment and labour force participation rates are time-invariant and equal across regions (Paci, 1997, p.612; see also Boldrin and Canova, 2001). Second, the analysis in terms of income per capita seems more appropriate if the main issue is to compare the economic standard of living across economies. When the focus is equity and how standard of living changes over time, the units of measurement should be the population rather than regions. This

[4]Club convergence is not necessarily in contrast to neoclassical growth models. Galor (1996) shows that multiplicity of steady state is consistent with the neoclassical paradigm when heterogeneity across individuals is allowed.

is relevant for an understanding of income differences of the European citizens, with the purpose of implementing distributive policies.

3 Data

Data are based on the Nomenclature of the Territorial Units for Statistics (NUTS) established by Eurostat. Functional criteria have also been adopted in the literature, as the Functional Urban Regions (FURs) which take into consideration the spatial sphere of socio-economic influence of any basic unit (Cheshire and Carbonaro, 1995; Magrini, 1999). However, data availability often requires to be attached to normative criteria for the selection of sub-national units. The common level of classification for the empirical analysis is the NUTS-2 level. However, for some countries this level does not correspond to the actual administrative organization of the countries. For instance, the NUTS-2 level is not adequate to an administrative level for UK and Denmark. Therefore, following Paci (1997), different digits of the NUTS classification have been selected for each country. The territorial units selected are the following 110 units: NUTS-0 (countries) for Denmark, Luxemburg and Ireland; NUTS-1 for Belgium (3 Régions), Germany (11 Länder), Netherland (4 Landsdelen) and United Kingdom (9 Government Office Regions & 3 Countries); NUTS-2 for Italy (20 Regioni), France (22 Régions), Spain (17 Comunitated Autonomas), Portugal (5 Comissaoes de coordenacao regional), Greece (13 Development Regions). The data-set, based on CRENoS data-bank [5], spans the period 1977–1996. The income variable is the GDP, calculated according to the 1979 European System of Integrated Economic Accounts (ESA 79). The GDP figures are at 1990 constant prices and are converted to Purchasing Power Standards (PPS) to control for differences in the cost of living[6].

In taking per capita GDP converted in PPS as a measure of standard of living, there are some concerns to be aware of. GDP is not equivalent to the final disposable income of private households in a given region. Therefore it can be considered only as an approximation of the capacity of individuals to acquire goods and services. Unfortunately, a time series of disposable income at regional level is not available. Only the full implementation of the new European System of Accounts (ESA 95) will provide regional household income figures. Furthermore, the GDP per inhabitant is affected by a commuting bias, the effect of commuters who live in one region and work in another. Regarding the different

[5] Crenos, *Center for North South Economic research*, is a research center interested in the analysis of development gaps at the national and regional level. Crenos databank can be downloaded from www.crenos.unica.it.

[6] Our data are spatially autocorrelated, reinforcing the fact that economic geography may help explain the economic growth process. For instance, the Moran's Index (with "gravity model" weighting scheme) ranges from 0.58 in the mid 1970s to 0.44 in the mid 1990s, while the Geary's c statistic ranges from 0.40 to 0.64. The positive spatial autocorrelation is statistically significant (the null hypothesis of no spatially autocorrelation is always rejected), but declines over time.

purchasing powers of the territorial units, only national PPS have been applied, since Eurostat does not estimate comparable regional price levels which would enable us to take into account regional differences in price levels within the same country.

As previously pointed out, weighing each region by its population, shifts the unit of observation from region to population, assuming equal distribution within each region. The selected 110 regional units are not equally sized in terms of population, with some remarkable differences. The variability across regions is notable, though not comparable with differences in world population distribution (see, e.g., Jones, 1997). For instance, in 1977 the population average is 2,843,000 ranging from 112,000 (Valle d'Aosta) to 17,073,000 (Nordrhein-Westfalen). The standard deviation is 2,681,000, corresponding to a relative standard deviation of 8.99 per cent. In 1996 the population average is 3,040,000 ranging from 119,000 (Valle d'Aosta) to 17,893,000 (Nordrhein-Westfalen). The standard deviation is 2,849,000, corresponding to a relative standard deviation of 8.94 per cent. Therefore, the population structure seems stable over the time period and eventual changes of the shape of the income distribution are not essentially due to changes in the population-based weights.

Since large movements in the distributions are not expected year-after-year, greater changes in the static distributions can be observed by analysing selected years, as representative of a specific time period. In particular, the year 1977 stands for mid 1970s, 1982 for early 1980s, 1987 for mid 1980s, 1992 for early 1990s and 1996 for mid 1990s. Distributions have been estimated in terms of logarithm of income.

4 Features of EU regional income densities

4.1 Kernel density estimation

For each year of the analysis, the densities are fitted by using a weighted kernel density estimator. Estimation should not abstract from the size of the different regions. The weights in the kernel estimator avoid giving excessive weight to regions with low populations and a modest weight to the densely populated regions. Formally, given a vector of n observations, $X_1, X_2, \cdots, X_n$, the estimator at point x is defined as:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \omega_i K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where h is the bandwidth that governs the degree of smoothing, ω_i the weight attached to the i-th observation, which reflects the different information of each observation about the underlying density. While the choice of the kernel function $K(\cdot)$ is considered in the literature a minor issue (Silverman, 1986), the selection of the bandwidth is crucial since the resulting estimated density is highly influenced by the smoothing parameter h. The consistently good performance over a wide range of smooth density shapes (Jones, Marron and Sheather, 1996) of the Sheather and Jones (SJ, 1991) *plug-in* procedure, suggests it as the method of selection of the optimal bandwidth.

The estimated densities reported in Figure 1 reveal several changes over this period[7]. The mass of the distribution moderately shifts towards the right (the kernel estimated mean value grows by 1.5 per cent over the period), while the level of dispersion exhibits a small reduction (the kernel estimated standard deviation decreases by 4.1 per cent). Perhaps, most markable is the change of the shapes of the distributions. While until the mid 1980s the densities are negatively skewed and clearly show at least two modes, one corresponding to low-income class and the other one to middle-income ranges, during the 1990s the distribution becomes more positively skewed and the smaller peak progressively collapses. This finding may suggest an improvement in economic conditions of the poorest regions and an inner process of catching up. One drawback of these figures is that the analysis is based only on a visual impression. In addition to this graphical approach, statistical tests are also helpful to reveal certain features of the shapes that otherwise may be undetected.

4.2 Multimodality and changing shape

It is well documented that the world distribution of output per capita across countries changed from a unimodal and close to log-normal distribution in the 1960s to a clear bimodal distribution in the 1980s and 1990s, indicating a group of "poor" and a group of "rich" countries. This observation is in evident contrast with the process of convergence, corresponds to the so-called twin-peakness and has been explained in terms of club convergence (Quah, 1997). Evidence for the distribution of per capita income across EU regions is much less clear. Quah (1996b) does not find evidence supporting a twin-peaks phenomenon over the 1980-1989 period (for a sample of 78 regions). Instead Magrini (1999), using similar empirical approaches, find such evidence for extended samples of European regions. However, the two modes in the European regional distributions are not as well separated as those detected in the cross-country distribution (Le Gallo, 2004).

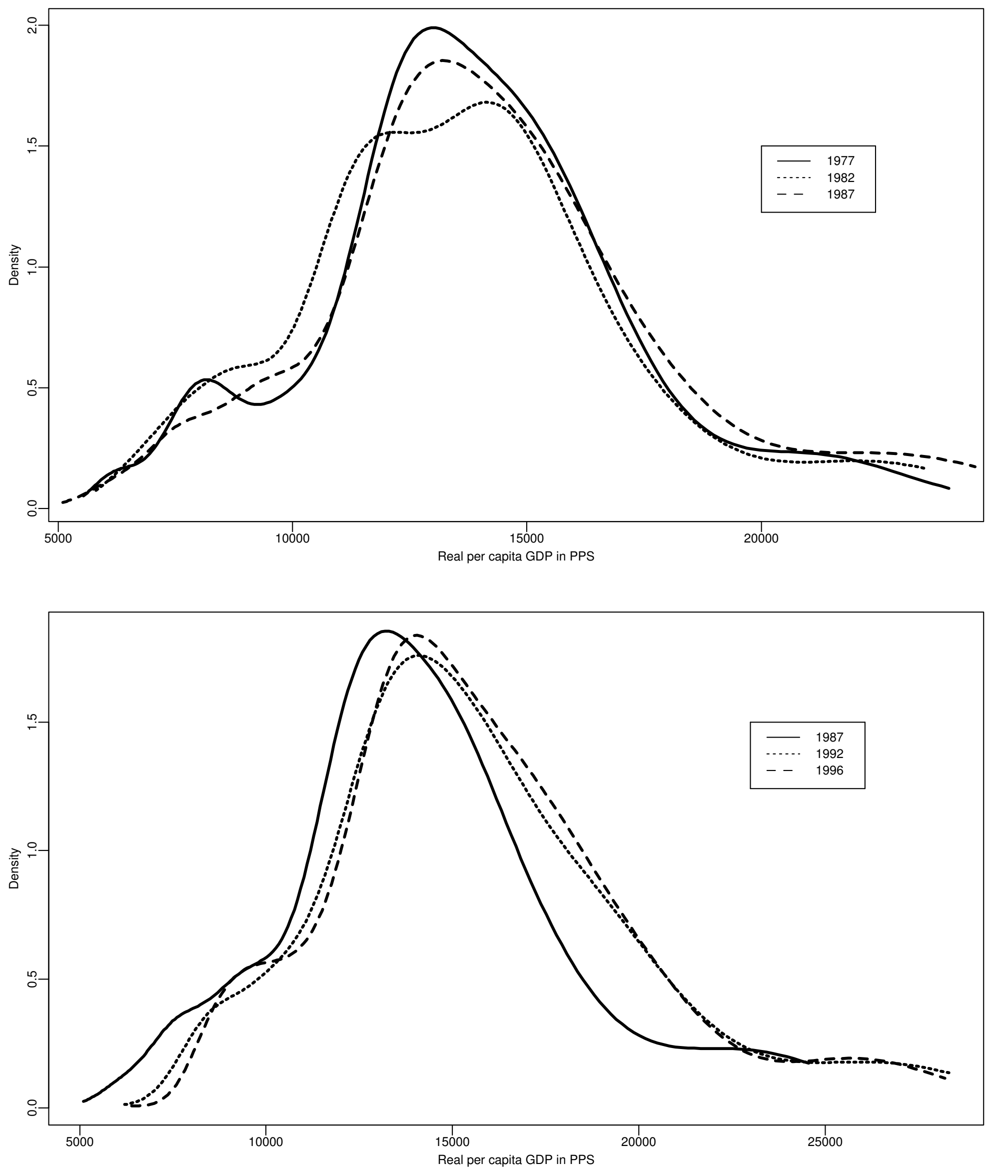
To statistically assess the presence of more than one mode[8] in the distributions, Silverman (1981) has introduced a test based on the "critical smoothed bandwidth" $h = h_m$ of the kernel density estimator, that is the smallest possible value of h producing a density with m modes.

The intuition of the test is that if the underlying density has $m + 1$ modes, a "large" value of h_m is expected because a considerable amount of smoothing is required to obtain a m-mode density from a $(m + 1)$-mode density. This suggests that h_m can be used as

[7] The estimated income densities are reported by transforming back the data to the original scale.

[8] A mode here is defined as a point at which the gradient of the density changes from positive to negative.

Figure 1: Weighted kernel estimates of EU regional income densities: the top panel shows the densities for the mid 1970s, early 1980s and mid 1980s, while the bottom panel the densities for the mid 1980s, early 1990s and mid 1990s.



a statistic to test the null hypothesis that $\widehat{f}(x)$ has m modes versus the alternative that $\widehat{f}(x)$ has more than m modes. A large value of h_m indicates more than m modes, thus rejecting the null. How "large" h_m should be, is assessed by the bootstrap test[9].

The test proceeds as follows. From the density, which has been rescaled[10] so as to have the same variance as the original sample, B samples with replacement are drawn. For each bootstrap sample $x^*(b) = (x_1^*, x_2^*, \cdots, x_n^*)$ the critical bandwidth $h_m^*(b)$ consistent with m-modality is computed. An estimate of the p-value associated with the corresponding critical bandwidth is: $\# \{h_m^*(b) > h_m\} / B$, where $b = 1, \cdots, B$. The null hypothesis of m modes in the density is rejected whenever the estimated p-value is smaller than standard levels of significance.

In Table 1 the critical bandwidths and the estimated p-values of the test with $B = 500$ replications are reported for each year of the analysis. The visual impression is in general confirmed, even though some differences can be noticed. In the mid 1970s and in the early 1980s the hypothesis of unimodality is clearly rejected, while the hypothesis of bimodality cannot be rejected. In the mid 1980s the erosion of the second peak in the left tail of the distribution makes the detection of multimodality less evident. In fact, the hypothesis that the distribution is unimodal cannot be rejected at the usual levels of significance but the p-value is just above 10 per cent. In the early 1990s results support the hypothesis of unimodality. In the mid 1990s, the detection of multimodality is more intriguing, since the presence of one, two and three modes is data-compatible (the p-values associated to h_1 and h_2, respectively 0.14 and 0.16, are just above the standard levels of significance). It should be noted that this test is conservative, that is the test is less likely to falsely reject the null hypothesis (Efron and Tibshirani, 1993).

As pointed out by several authors (e.g. Bianchi, 1997), detection of multimodality can be suggestive of several income groups, each referring to certain economic and geographical characteristics. These groups can be estimated by a mixture of distributions with different mixing proportions, although there is no guarantee of biunivocal correspondence between the number of modes and the number of unimodal components in the mixture (Izenman and Sommer, 1988), except when the modes are "well separated".

Although the number of modes of the distribution has important implications in terms of eventual sub-groups, finding the same number does not necessarily imply time invariance of the shapes. Observed changes in shape dynamics can be due to random variation or to actual features of the underlying distributions. To assess whether the differences between

[9] The critical bandwidth is uniquely defined if the Gaussian kernel is used because in that case the number of modes of the estimated density is non-increasing in h (see Silverman, 1981).

[10] This small adjustment rescales in order to have the same variance as the original sample since the kernel estimation in (1) artificially increases the variance of the estimate (Efron and Tibshirani, 1993). This is, rather than sampling with replacement from the data, a sampling from a smooth estimate of the population. For this reason it is called smoothed bootstrap.

Table 1: Test for m-modality in the distributions: critical bandwidths and relative p-values.

Time period	H_0: $f(x)$ has m modes	h_m	p-value
mid 1970s	$m = 1$	0.104	0.01
	$m = 2$	0.063	0.38
	$m = 3$	0.051	0.17
early 1980s	$m = 1$	0.093	0.04
	$m = 2$	0.070	0.43
	$m = 3$	0.065	0.51
mid 1980s	$m = 1$	0.092	0.12
	$m = 2$	0.076	0.48
	$m = 3$	0.071	0.31
early 1990s	$m = 1$	0.094	0.38
	$m = 2$	0.077	0.51
	$m = 3$	0.064	0.73
mid 1990s	$m = 1$	0.088	0.14
	$m = 2$	0.072	0.16
	$m = 3$	0.064	0.68

estimates reflect systematic differences in the densities structure, a nonparametric test is carried out. To restrict the analysis to relative effects, weighted regional per capita GDP are normalised by the corresponding European mean for each year. In this way, only significant changes in the mass of the density are detected regardless of shifts of the location of the whole distribution. The problem of testing $H_0 : f(x) = g(x)$ against $H_1 : f(x) \neq g(x)$ can be carried out by the most natural test statistic, the integrated squared error:

$$\widehat{I} = \int_x [\hat{f}(x) - \hat{g}(x)]^2 dx$$

where $\hat{f}(x)$ and $\hat{g}(x)$ are the non parametric estimates of the two unknown distributions $f(x)$ and $g(x)$. For a better comparison, the same amount of smoothing should be applied (Marron and Schmitz, 1992). The geometric mean of the individual optimal bandwidths is considered as the smoothing parameter for comparison of the density estimates. The value of the integral can be computed analytically (Li, 1996) or can be approximated by numerical integration. Under commonly used assumptions, applying a central limit theorem for degenerate U statistics, Li (1996) shows that, under H_0, $h \longrightarrow 0$, and $nh \longrightarrow \infty$, a transformation of $\widehat{I}$ converges to a standard normal distribution. This result is valid where data are either independent or dependent. However, for small samples the normal distribution may not provide an accurate approximation of the test statistics under the null hypothesis. Consequently, the distribution of the statistic is evaluated by an empirical procedure. If the null hypothesis is correct, the two groups of data are generated from the same underlying density function. In this statement, as Bowman and Azzalini (1997) suggest, all the $2n$ observations from both groups are combined together. A sample of size n, which represent the first group, is randomly taken without replacement from the pooled data. The remaining n observations constitute the second group. The integrated squared difference is numerically approximated over a finely spaced grid and the process is repeated $B = 500$ times. The empirical significance of the statistic is given by the proportion of simulated values greater than the value of $\widehat{I}$ calculated in the original data.

The results of comparison between the cross-regional distributions over time are reported in Table 2. In general, the test gives evidence against immobility, indicating that something in the EU continues to evolve. What emerges is a significant difference between the distributions in the 1990s with respect to the preceding distributions. In fact, all the pairwise comparison between the 1990s distributions and those in the preceding years show clear evidence against the null hypothesis. Instead, not very significant changes are detected in the shapes of relative income between the early 1990s and the mid 1990s. Regarding the distributions before the 1990s, no equally clear conclusions can be made. Although not formally statistically significant, the empirical p-values indicate a "borderline" evidence when comparing the mid 1980s distribution against the previous ones.

Table 2: Test for changing shape in the relative regional per capita income distribution: the integrated squared differences and relative p- values.

Year	mid 1970s	early 1980s	mid 1980s	early 1990s	mid 1990s
mid 1970s	-	$I = 2.14$ $[p = 0.01]$	$I = 1.04$ $[p = 0.12]$	$I = 3.47$ $[p = 0.00]$	$I = 3.62$ $[p = 0.00]$
early 1980s		-	$I = 0.98$ $[p = 0.16]$	$I = 1.33$ $[p = 0.06]$	$I = 1.47$ $[p = 0.07]$
mid 1980s			-	$I = 1.40$ $[p = 0.03]$	$I = 1.62$ $[p = 0.05]$
early 1990s				-	$I = 0.78$ $[p = 0.26]$

5 Modelling the EU income distribution by a finite mixture

5.1 Mixture models

Mixture models can be viewed as a semi-parametric alternative to the non-parametric densities[11], especially when the non parametric density exhibits more than one mode, and provide greater flexibility and precision in modelling the underlying distributions of

[11] In particular, fitting an unknown density by a mixture of $G = n$ components in equal proportions $1/n$, where n is the size of the observed sample, is equivalent to a non-parametric kernel estimation. Therefore, mixture models can be viewed as a compromise between parametric models, represented by a single parametric family ($G = 1$) and a non-parametric model, represented (in the case of $G = n$) by the kernel density estimator.

sample data.

As any continuous distribution can be well approximated by a normal mixture (Marron and Wand, 1992), the log-incomes are considered as coming from a mixture of G normal distributions in proportions $\pi_1, \cdots, \pi_G$. One of the main advantages in using mixture models is that, once a model is generated, conditional probabilities τ_{ig} which represents the probability that region i with income x_i comes from the gth component of the mixture can be computed for each region.

Formally, the probability density function under a G-component mixture model is defined as:

$$f(x_i, \mathbf{\Psi}) = \sum_{g=1}^{G} \pi_g f_g(x_i, \theta_g) \tag{2}$$

where the vector $\mathbf{\Psi} = (\pi_1, \cdots, \pi_{G-1}, \mathbf{\Theta}')'$ contains all the unknown parameters in the mixture model; π_g, $g = 1, \cdots, G$ represent the mixing proportions and the vector $\mathbf{\Theta}$ contains all the parameters $(\theta_1, \cdots, \theta_G)$ known *a priori* to be distinct; $f_g\,(x_i, \theta_g)$ denotes the values of the univariate density specified by the parameter vector θ_g; in this case $\theta_g = \left(\mu_g, \sigma_g^2\right)$ denotes the mean and the variance of the univariate normal component g. The mixing proportions $\pi_1, \cdots, \pi_G$, that are nonnegative and sum to one, give the prior probability that a region belongs to the gth component of the mixture, representing an endogenous parameter which determines the relative importance of each component in the mixture. The fitting of the mixture model provides a probabilistic clustering of the n regions in terms of their estimated *ex post* probabilities of membership of the individual G components of the mixture of distributions. The posterior or conditional probability τ_{ig} is given by:

$$\tau_{ig} = \texttt{Prob}\{C(i) = g \mid (x_i; \mathbf{\Psi})\} = \frac{\pi_g f_g(x_i)}{\sum_{h=1}^{G} \pi_h f_h(x_i)} \tag{3}$$

where $C(i)$ indicates the component to which region i belongs.

Different methods can been used to estimate the parameters, but the iterative fitting by maximum likelihood (ML) via the expectation-maximization (EM) algorithm, with the assignment of the component for each observation regarded as the missing information (Dempster *et al.*, 1977), seems to be superior to other procedures, as reported in McLachlan and Peel (2000). In this analysis, a variant of the EM algorithm is applied to enclose the population-based weights attached to each region. Starting from a weighted log-likelihood, the E step comprises estimation of the conditional probabilities, while the M step estimation of the densities f_g and the marginal probabilities π_g. The marginal probabilities π_g are estimated as the weighted means of the region-specific probabilities τ_{ig}, with $\boldsymbol{\pi}$=$(\pi_1, \cdots, \pi_G)$ and $\mathbf{f} = (f_1, \cdots, f_G)$ replaced by their current estimates. The estimation of the normal component f_g entails computing means and variances weighted by regional population. The E and M steps are applied iteratively and repeated as necessary. Each iteration increases the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function. However, the likelihood function for mixture models usually has several local maxima and, in mixtures of univariate normal components with unequal variances, it is unbounded and the global maximizer does not exist. The largest of the local maxima located is obtained applying a range of starting values for the EM algorithm. Here, random starts and k-means (crisp and fuzzy) clustering-based starting values are selected for initialization.

5.2 The number of components of the mixture

The choice of the number G of component densities compatible with the data is a difficult problem that has not been completely solved (Aitkin and Wilsonin, 1980; Richardson and Green, 1997; McLachlan and Peel, 2000). Starting from *a priori* information, Paap and van Dijk (1998) assume a finite mixture of two components in modelling the world distribution of the GDP per capita over the period 1960-1989. Since for the European income distribution there is no information *a priori* regarding the number of components in the mixture model, the selection of G has to be assessed. A natural way to test the null hypothesis $H_0 : G = G_0$ versus $H_1 : G = G_1$, for some $G_1 > G_0$, could be the use of the likelihood ratio (LR) test statistic $-2\log\lambda$:

$$-2\log\lambda = 2\{\log L(\hat{\boldsymbol{\Psi}}_1) - \log L(\hat{\boldsymbol{\Psi}}_0)\}. \tag{4}$$

Unfortunately, in mixture models, the regularity conditions for $-2\log\lambda$ do not hold and the asymptotic distribution of the test statistics is not a chi-squared (Quinn, McLachlan and Hjort, 1987). Consequently, the null distribution of $-2\log\lambda$ can be estimated by a bootstrap procedure (McLachlan, 1987).

The test proceeds as follows. Under the null hypothesis of G_0 components, B bootstrap samples are generated parametrically. Each bootstrap sample is fitted with G_0 and G_1 mixture model and the quantity $-2\log\lambda$ is computed. The null distribution of $-2\log\lambda$ can be finally estimated from B replicated values of $-2\log\lambda$. Therefore, the estimate of the achieved significance level p corresponding to the value $\widehat{\theta}^* = -2\log\lambda$ evaluated from the original sample, is:

$$p - \text{value} \simeq 1 - \frac{r}{B+1}, \tag{5}$$

where r is the position of $\widehat{\theta}^*$ in the sorted list of the replications. It should be reminded (see also Section 4.2) that when the mixture model is used in a clustering context, the components reflect distinct groups in the population, which cannot necessarily correspond to the number of modes detected in the distribution. For instance, even if the number

Table 3: The choice of the number of components

1977	AIC	BIC	$\widehat{\theta}^*$	p-val	1982	AIC	BIC	$\widehat{\theta}^*$	p-val	1987	AIC	BIC	$\widehat{\theta}^*$	p-val
G=1	24.4	29.8	-	-	G=1	33.3	38.7	-	-	G=1	35.7	41.1	-	-
G=2	13.8	27.3	16.5	0.00	G=2	29.0	42.5	10.2	0.09	G=2	31.4	44.9	10.3	0.08
G=3	15.0	36.6	4.8	0.52	G=3	31.9	53.5	3.12	0.62	G=3	32.3	53.9	5.12	0.46
G=4	21.0	50.7	0.02	0.96	G=4	33.0	62.7	4.9	0.47	G=4	33.1	62.8	5.18	0.42
1992	AIC	BIC	$\widehat{\theta}^*$	p-val	1996	AIC	BIC	$\widehat{\theta}^*$	p-val					
G=1	24.3	29.7	-	-	G=1	20.6	25.9	-	-					
G=2	26.1	39.6	4.2	0.64	G=2	20.5	34.0	6.0	0.17					
G=3	28.1	49.8	3.9	0.67	G=3	18.6	40.1	8.0	0.09					
G=4	34.0	63.7	0.02	0.94	G=4	20.9	50.6	3.7	0.58					

of modes can be suggestive of the number of separate underlying income distributions, a mixture distribution can be also unimodal when the component are not sufficiently far apart. On the other hand, bimodality does not necessary imply that the data have been sampled from a two-component mixture distribution.

Table 3 reports, for each year of the analysis, the value of the test statistic $\widehat{\theta}^*$ and the corresponding p-values for testing the hypothesis of G components against the alternative $G+1$, for $G = 1$ to $G = 3$. The values of the Akaike and Bayesian Information Criteria are also listed in the table. The number of components for each year is chosen according to the p-values of the bootstrap test. As reported in the table, the results of the bootstrap test are consistent with the Akaike criterion, while the Bayesian criterion does not always agree with the bootstrap test. The hypothesis of bimodality in the mid 1970s and early 1980s, detected by the Silverman's test, is corroborated by the presence of two components in the mixture. In fact, the LR test supports the hypothesis of two components in the mixture in both 1977 and 1982. In the mid 1980s, due to the erosion of the peak in the left tail of the distribution, the multimodality test is not able to capture the bimodality. Instead, the LR test supports the evidence of two components in the distribution, since it rejects the hypothesis of one component at 8 per cent level of significance. In 1992 the distribution is unambiguously unimodal. In fact, the hypothesis of one component is clearly not rejected, consistently with the multimodality test. In the last year of the analysis, the question of whether the distribution is unimodal is less trivial. The hypothesis of one component can not be rejected at the 17 per cent significance versus the alternative of two components. However, when we test $G = 2$ components against $G = 3$ the null hypothesis of two components is rejected at a 10 level of significance, while the hypothesis of three components versus $G = 4$ can not be rejected. Therefore, in the mid 1990s, the per capita GDP distribution in EU could be estimated by a single component but a

Table 4: Means, standard deviations and mixing proportions of the fitted mixture models.

	Mean			Standard deviation			Mixing proportion		
years	g_1	g_2	g_3	g_1	g_2	g_3	g_1	g_2	g_3
mid 1970s	7931	14031	-	1138	2597	-	0.145	0.855	-
early 1980s	7821	13602	-	1059	3013	-	0.118	0.882	-
mid 1980s	7821	14183	-	1314	3137	-	0.110	0.890	-
early 1990s	14854	-	-	4048	-	-	1	-	-
mid 1990s	9369	15217	26179	901	2786	1276	0.114	0.846	0.040

mixture of three components seems to better approximate the empirical distribution. For each year, Table 4 reports the values of the means, standard deviations and the estimated marginal probabilities of the components of the fitted mixture model. For convenience, we report the means and the standard deviations of the distribution of the real income in levels, obtained by back-transformation.

For the mid 1970s, the fitted normal mixture density consists of two components, which are fairly wide apart. The first component is characterized by a group of poor regions, with mean value equal to 7931 PPS, representing 14.5 per cent of the whole populations, and a group of relatively richer regions, with mean equal to 14031 PPS, representing the remaining 85.5 per cent of the population. Looking at the conditional probabilities[12] $\widehat{\tau}_{ig}$, most of the regions can be allocated to a component with a high degree of certainty. Portuguese regions except for Lisboa and Vale do Tejo, eleven Greek regions, five Spanish regions (Canarias, Extremadura, Andalucia, Galicia and Castilla-la Mancha), four Southern Italian regions (Calabria, Sicilia, Campania and Molise) and Ireland have an ex post conditional probability to belong to the first component of the mixture greater than 85 per cent, while (West) Germany, Benelux, France and UK (except Northern Ireland) regions all belong to the "rich" group with an *ex post* conditional probability larger than 90 per cent. In contrast, only for two regions (Puglia and Murcia), the mixture component is uncertain, as the conditional probabilities are in the range 0.35–0.66.

The distribution for the early 1980s is still well approximated by two components. The first component still represents a group of poor regions, but it accounts for only the 11.8 per cent of the population. This reduction is essentially due to the decrease of τ_{i1} in some regions, especially the Italian ones. The second component, with mean equal to 13602 PPS, roughly corresponds to the previous "rich" group.

A two-component mixture is also compatible with the mid 1980s data. The mixing proportions of the components have not significantly changed. However, the two compo-

[12] Due to space constraint, the posterior probabilities for each region and each year are not reported here, but they are available upon request.

nents partially overlap as the allocation of regions to groups entails more uncertainty. In fact, 9 per cent of the regions are assigned to a component with a relatively low probability, ranging from 0.35 to 0.66.

Results obtained for the early 1990s show a tendency of the distribution to be generated by a single component, with mean equal to 14854 PPS and standard deviation equal to 4048. This suggests that a process of convergence occurs and it is no longer possible to detect a well defined group of poor and a group or rich regions from the data.

Also in the mid 1990s, the hypothesis of one component is not rejected from the data at the 10 per cent level of significance. However, in Table 4 we report the results for a fitted model of three components as the LR test also supports this hypothesis. The model fit departs from the previous patterns. The dominant component accounts for 84.6 per cent of the population. Its mean is 2 per cent above the EU-12 average and its standard deviation 28 per cent smaller. The first and third components, which substantially improve the model fit, share the remainder of the population, respectively 11.4 per cent and 4 per cent. The first component still represents a group of poor and homogeneous regions, but its mean, 9369 PPS, is 20 per cent higher than the respective value of 1987. A small cluster of regions, characterised by very high per capita GDP are generated by a third component with mean 26179 PPS. Administrative and service regions (Hamburg, Bruxelles, Île de France and Luxembourg), have a posterior probability to belong to this component exceeding 0.8.

In summary, the European regional distributions in the 1970s and 1980s exhibit a polarization in two disjointed clusters: a group of poor regions and a group of wealthier regions. In the early 1990s there is evidence of a process of convergence of the regions toward a single dominant group. However, in the mid 1990s, also a mixture of three components fits the data well. The dominant component, which attracts most of the regions into its domain, is similar to its counterpart in the two component mixture (for the mid 1970s, early 1980s and mid 1980s) and in the one component model (for the early 1990s). The first component still represents a group of poor regions, but its distance from the dominant component is reduced, indicating a process of "catching up" of the poorest regions. Therefore, our empirical evidence shows that the process of economic integration may have contributed to equalising per capita income across regions. On the other hand, a small group of very rich regions in the upper tail of the distribution seems well captured by a separate component, indicating a tendency for the richest regions to diverge. Going back to the data, in fact, the ratio of the 95 percentile to the median increased from 1.42 in the mid 1970's to 1.71 in the mid 1990's. As pointed out in the data description section, per capita income level of city-regions may be overestimated by a commuting effect, and symmetrically that of neighbouring regions may be understated. Another reason that may induce overestimation is the adoption of a national PPS deflator, which presumably underestimate the actual PPS in the city-regions. Despite these concerns, it may be too simplistic to attribute the robust growth of the real GDP experienced by some rich regions only to measurement inaccuracy. Consequently, more effort should be devoted to establish the economic reasons why certain "core" regions have concentrated scale-intensive service activities, widening the income gap with the other regions.

5.3 Goodness of fit of the mixture

To visually show the quality of the fit, in Figures 2–6, the kernel density estimation, the fitted mixture and their components are reported for each year of the analysis. As revealed in the pictures, the detected mixture models seem to fit quite well to the data at hand. The graphical approach is supported by a kernel-density-based statistical test to investigate whether or not a finite mixture model is compatible with the data. The availability of kernel density estimates consents to make the comparison on a density scale rather than from a probability plot or empirical distribution function, as many traditional procedures. Specifically, the goodness of fit of the mixture model is assessed by a measure of global distance between the mixture and the kernel, $J = \int_x [f(x) - f(x; \boldsymbol{\Psi}_0)]^2 dx$, which can be estimated by $\widehat{J} = \int_x [\widehat{f}(x) - f(x; \widehat{\boldsymbol{\Psi}}_0)]^2 dx$, where $\widehat{f}(x)$ is the kernel estimation of the unknown density and $f(x; \widehat{\boldsymbol{\Psi}}_0)$ is the estimated mixture model. As argued in Fan (1994), different tests can be carried out according to the amount of smoothing applied to the data. Following Bowman (1992), to remove the bias produced by the smoothing, the comparison at each point x is made between the kernel density and the kernel-smoothed parametric estimate under the assumption of a mixture of normal components. Formally, the test contrasts the density estimate with its estimated mean under the assumption of mixture, that is:

$$\widehat{J^*} = \int_x \left[\widehat{f}(x) - \widehat{E}\widehat{f}(x)\right]^2 dx. \tag{6}$$

The estimated mean $\widehat{E}\widehat{f}(x)$ is a convolution of a kernel with a mixture of normal components $\widehat{E}\widehat{f}(x) = K_h * \widehat{f}(x) = \frac{1}{h} \int_u K\left(\frac{x-u}{h}\right) f(u; \widehat{\boldsymbol{\Psi}}_0) du$. When the kernel function is Gaussian, the convolution collapses into a mixture of normal densities with means equal to μ_g and variances equal to $(\sigma_g^2 + h^2)$. Under the null hypothesis of correct parametric specification, Fan (1994) proves that a center-free test statistic based on $\widehat{J^*}$ is asymptotically normally distributed (see also Fan and Ullah, 1999). In finite samples, however, the normal distribution may not provide an accurate approximation to the distribution of the test statistic. Indeed, obtaining a better approximation to the distribution of the test statistic motivates the use of a parametric bootstrap procedure, as proposed in Fan (1995). In this case, the bootstrap samples are drawn from the mixture distribution $f(x; \widehat{\boldsymbol{\Psi}}_0)$, where the parameters of the mixture, $\widehat{\boldsymbol{\Psi}}_0$, are estimated using the original sample. Table 5 records, for

Table 5: Test for goodness of fit of a mixture of normals: test statistic based on kernel methods and relative p-values.

Year	Number of components	$\widehat{J}^*$	p-value
mid 1970s	2	0.023	0.534
early 1980s	2	0.113	0.248
mid 1980s	2	0.037	0.326
early 1990s	1	0.133	0.234
mid 1990s	3	0.044	0.512

each year of the analysis, the statistic $\widehat{J^*}$ and its corresponding p-value calculated from $B = 500$ simulations under the null hypothesis of a mixture of G normal components. The number of components in the mixture is established according to the LR test previously described (*cfr.* Table 3).

The results reported in Table 5 confirm the graphical evidence that the selected mixture models do fit well, as the empirical p-values of all the test statistics are largely above the usual levels of significance.

6 The distribution dynamics across EU regions

6.1 Stochastic kernel

Previous sections focused on cross-sectional distribution of real GDP per capita, without explaining the movements of the regions from one period to another. The intra-distributional dynamics between groups of regions complete the analysis since would lend insights into any interregional patterns of economic interaction. One of the statistical tool of the intra-distribution mobility analysis is the quantification of the probabilities of transition between quantilies or states[13], as suggested by Quah (1993). This approach computes standard transition probability matrices, which contain the probabilities of moving between different discrete states, usually under a first-order Markov process.

The transition probabilities are very sensitive to the necessarily arbitrary grouping from discretisation and the Markov property itself can be destroyed from inappropriate choice of discrete cells (see Bartholomew, 1981). Several improvements have been

[13] Also the mixture models could be interpreted in a dynamic context, as the switching from one group to another one depends on the relative movement of a region in the distribution with respect to other regions. However, this evaluation is possible when the number and the features of the components do not change over time. But it is not straightforward when the fitted mixture models present different characteristics, like in this analysis.

Figure 2: Kernel density estimation and the two–component mixture model fit in the mid 1970s.

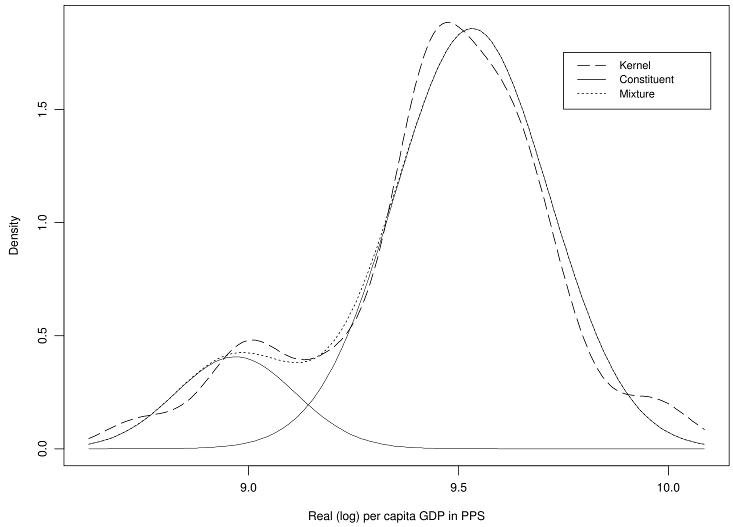


Figure 3: Kernel density estimation and the two–component mixture model fit in the early 1980s.

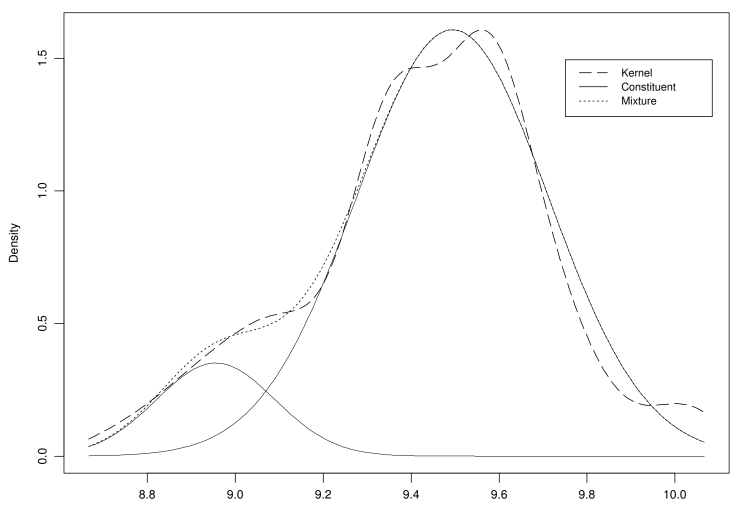


Figure 4: Kernel density estimation and the two–component mixture model fit in the mid 1980s.

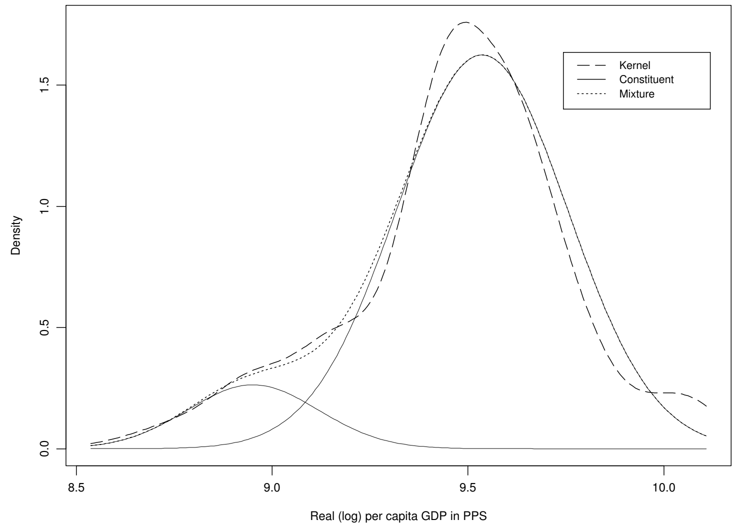


Figure 5: Kernel density estimation and the one–component mixture model fit in the early 1990s.

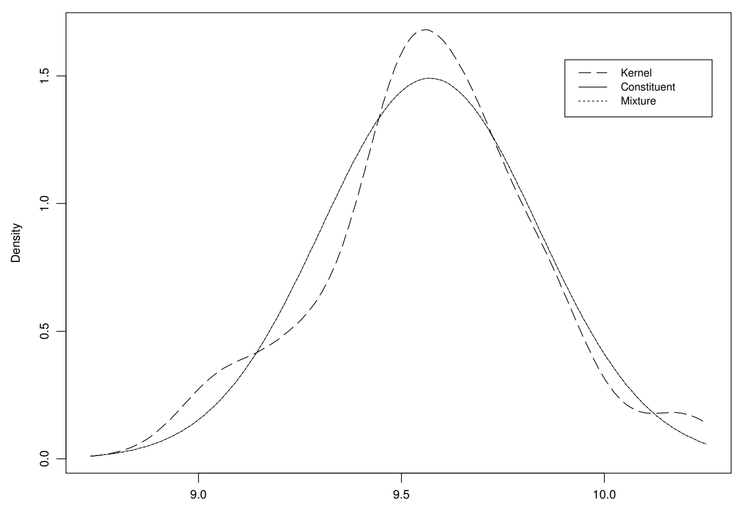
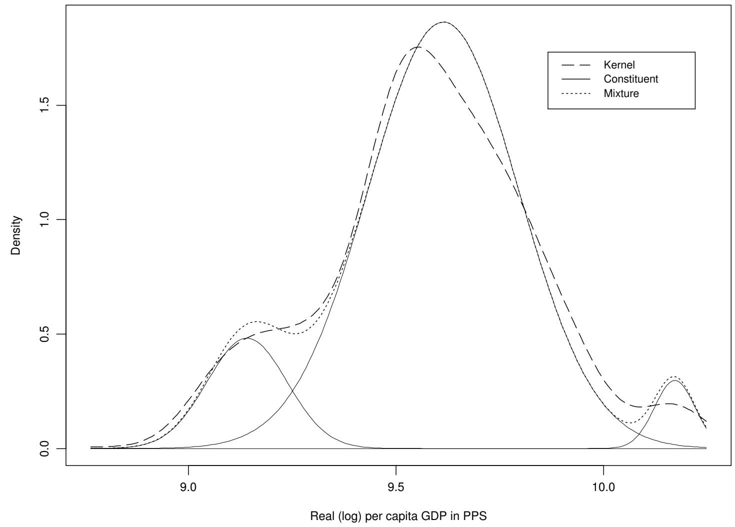


Figure 6: Kernel density estimation and the three–component mixture model fit in the mid 1990s.



suggested. Fingleton (1997) shows that modelling the transition probabilities may help trascend the particular discretisation. Bulli (2001) explores in some detail the difficulties involved in choosing the appropriate discretisation for the income distribution, trying to find "optimal" ways to discretise the income space preserving the Markov property. Reichlin (1999) states that the long-run implications of the transition dynamics are particularly sensitive to the discretisation scheme [14]. Quah (1996c; 1997; 2001) invokes a continuous state space approach for income variable, introducing the transition function or stochastic kernel, the continuous equivalent of the transition probability matrix. In the continuous income state-space, the first-order transition function can be described by a conditional density function[15]. If the transition mechanism is time invariant, the stochastic kernel depends only on the lag τ and can be written as $f_\tau(y|x)$, giving, conditional on income x the τ-period-ahead density of y. The conditional density function is obtained dividing the joint distribution by the marginal one:

$$f_\tau(y|x) = \frac{f_{t,t+\tau}(x,y)}{f_t(x)}. \tag{7}$$

The joint distribution is estimated by a two-dimensional kernel evaluated on a square grid:

$$\hat{f}_{t+\tau}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} \omega_i K\left(\frac{x-X_i}{h_x}, \frac{y-Y_i}{h_y}\right), \tag{8}$$

[14] On the other hand, to weaken the problem of measurement errors, one can make the income variable categorical.

[15] A far more general formalization can be found in Quah (1997).

where K is a bivariate kernel with (double) integral equal to one; $z_i = (x_i, y_i)$ is the ith observation, ω_i the attached population-based weight. Similar to the univariate kernel, the bandwidths h_x and h_y establish the degree of smoothness of the estimates. The kernel function we use in the estimation of (8) is the bivariate Gaussian kernel. The bandwidths are chosen according to the Sheather and Jones *plug-in* method. The stochastic kernel is estimated for a 5-year transition period, setting $\tau = 5$. To remove the common growth and business cycle effects, the income level in each region is divided by the (population weighted) average of the figures of the 110 NUTS, obtaining a measure of relative income. The relative income data, observed from 1977 to 1996 give observations on fifteen 5-year transitions for each region to estimate $f_5(y|x)$. Given the joint distribution, the marginal distribution of x is derived by numerically integrating the joint distribution with respect to y.

If the transition function $f_\tau(y|x)$ describes the evolution of the income distribution over time, the relationship between the cross-region income distribution at time t, $f_t(x)$, and the income distribution τ periods later, $f_{t+\tau}(y)$, can be written as:

$$f_{t+\tau}(y) = \int_y f_\tau(y|x)\, f_t(x)\; dx. \tag{9}$$

Taking expression (9) to the limit as $t \longrightarrow \infty$ provides the ergodic (long-run) density implied by the transition function. It can be found as the solution to (Johnson, 2000):

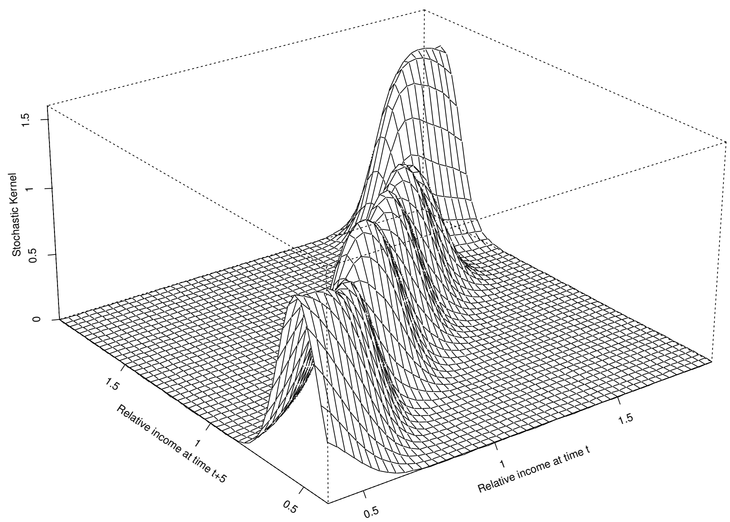
$$f_\infty(y) = \int_y f_\tau(y|x)\, f_\infty(x)\; dx. \tag{10}$$

6.2 Income mobility within regional distribution

The stochastic kernel provides insights about mobility and persistence in the estimated distribution of per capita income across European regions. The estimated three dimensional stochastic kernel is reported in Figure 7. Standing at any point on the period t axis, and then look in a straight line, parallel to the period $t+5$ axis, one obtains a probability density. The 45-degree diagonal indicates persistence properties and when most of the graph is concentrated along this line, then the elements in the distribution remain where they started. As evident from Figure 7, a large portion of the probability mass remains clustered along the diagonal, and most of the peaks lie along the 45 degree diagonal indicating low mobility and modest change in the regional cross-section distribution.

All the points on the stochastic kernel at a certain height are connected by one contour. Figure 8 depicts the relative contour plot of Figure 7. The contour plots indicate different levels of iso-probs, the outer ones indicating a lower probability. If the ridge of the stochastic kernel has, as its projection onto the base of the graph, the positive slope diagonal, then there is high persistence. The degree of persistence is higher when the width

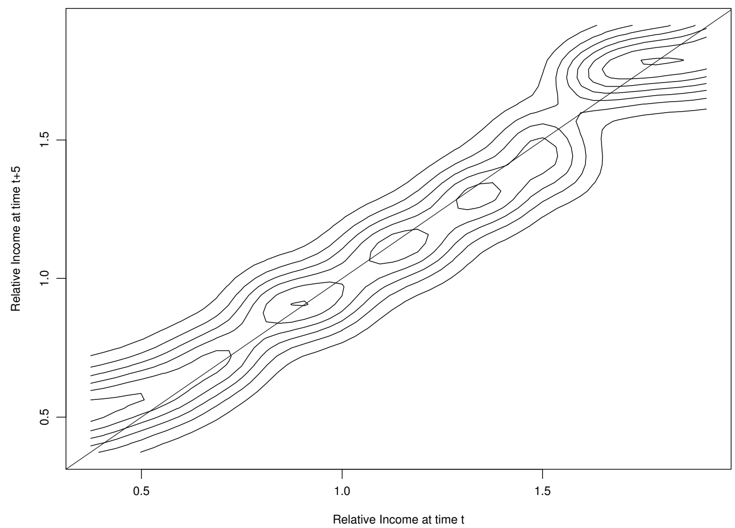
Figure 7: Relative income dynamics across 110 European regions: the estimated $f_5(y|x)$



of the iso-probs around it is lower. Instead, movements of the ridge of the kernel imply a certain degree of mobility. The contour plot in Figure 8 reveals some other features of the distributional change besides persistence already detected in Figure 7. The presence of some peaks off the diagonal emphasises a slight clockwise rotation. In fact, there is one peak above the 45 degree line corresponding to relative income values of 0.5, while there are some peaks below the diagonal corresponding to regions with income above average. This means that, even if the dominant feature is a low degree of mobility, regions with very low incomes tend to increase their relative position over the 5-year transition horizon, indicating a process of catching up of the poorest regions with the richer ones. Yet, the group of regions with income from 1.1 to 1.4 times the average income loses out in relative terms, becoming relatively poorer. In contrast, the position of a very rich group around 1.8 times the average remains unchanged, as indicated by the upper end peak lying along the diagonal. Evidence of Figure 7 and 8 are corroborated by the ergodic density function obtained solving equation (10), which describes the income distribution across European regions achieved in the long-run under the hypotheses previously described.

The stationary distribution across the 110 regions, plotted in Figure 9, is distinctively bimodal. The higher peak represents regions clustered just below the average value of one, while a group of relatively very rich regions gather around the smaller peak. The bimodality of the ergodic distribution is confirmed by a two-component mixture model (see Figure 9). The dominant component, which accounts for 92 per cent, comprises regions with mean equal to 0.96. This imply the emergence of a distinctive middle-income club, suggesting that in the long-run there is no development trap into which the poorest regions

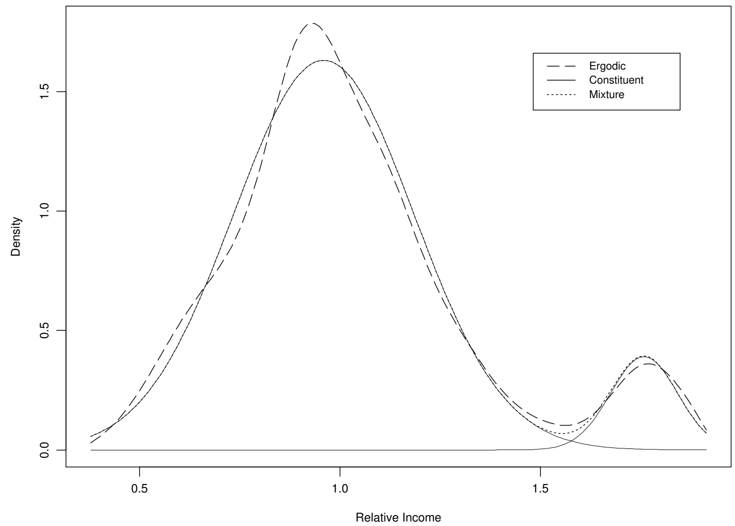
Figure 8: Relative income dynamics across 110 European regions: contour plot of the estimated $f_5(y|x)$



will be permanently condemned. In contrast, the mean income of the second component, accounting for 8 per cent of the entire population, is 1.8 times the average. This component identifies a small club of high-income regions.

The shape of the ergodic distribution obtained with our data differs from the unimodal ergodic solution drawn in Quah (1996b). With respect to our data set, his results are based on purchasing power standardised per capita income in a reduced sample of 78 regions (excluding Greece, Portugal and France) over the 1980-1989 period. Our evidence is also different from López-Bazo *et al.* (1999). Their ergodic distribution, estimated on a sample of 143 regions over the period 1980-1992, shows high probability mass in the lower income state, indicating a lack of convergence of the poorest regions consistent with multiple steady-states across EU regions. Also Fingleton (1999), relating loglinear models to Markov chain analysis, reaches similar conclusions in a data set of 178 NUTS-2 regions' manufacturing productivity for the period 1975-1995. Our evidence is coherent to Magrini (1999) stationary distribution calculated over the period 1979-1990. His evidence supports a process of slow convergence of NUTS-2 regions towards the same income class with the exception of a small group of "growth leaders", that is a group of regions growing away from the rest of the other European regions. It should be pointed out that our empirical results also reflect the effect of population-based weights attached to each regional income.

Figure 9: Estimated ergodic density and fitted mixture model



7 Concluding remarks

Our analysis of European regional income disparities has emphasised the relevance of non parametric and semiparametric models in detecting some features of the entire distribution that otherwise would go unnoticed. Dispersion of incomes between regions is measured taking into account their size. The population-weighted regional distributions have been initially described by kernel density estimates, which reveal, until the 1990s, a bimodal distribution, and since the early 1990s a trend towards unimodality, although a small peak in the right tail of the distribution seems to emerge. The visual inspection of the kernel estimates has been reinforced by introducing a Silverman-based multimodality test to judge whether the modes are spurious or not. To discriminate if the distributions are close, a non parametric changing shape test has been carried out. Our findings show that cross-section distributions are characterised by a multimodality structure until the 1990s, while the hypothesis of one mode can not be rejected in the 1990s as well as the presence of three modes in the mid 1990s. The shape of the distributions changes over time, indicating that something in the EU continues to evolve. A visible disadvantage of the kernel estimation is that the estimated density can not be directly related to any subpopulation. Since mixture models are better suited for inferences about distinct subpopulations, we introduced mixture modelling as a complementary to kernel densities estimation. To determine the number of components to retain in the mixture models, we have used a LR test, bootstrapping its empirical distribution. The goodness of fit of the estimated mixture models has been evaluated by a kernel-density-based statistical test. Empirical results show that the distributions of the population-weighted regional income are well approximated by a two component mixture in the 1970s and in the 1980s, whose components represent two well separated clusters of citizens belonging to poor and rich regions. Over time, these two clusters tend to merge into a single group, supporting the idea of a process closing of per capita income gaps, especially after the 1988 reform of the Structural Funds which significantly increased the budget for regional support policies. However, in the mid 1990s, a small group of very rich regions seems to form a separate component.

To model the intra-distribution dynamics, a stochastic kernel has been estimated under the hypotheses of time invariance and first order evolution. The concentration of the ridges of the three dimensional stochastic kernel and the disposition of the probability mass along the main diagonal in the two-dimensional countour plot imply a certain degree of stability and persistence of the distributions. However, a clockwise turning of the peaks of the stochastic kernel leads to a long run distribution different from the current distributions. It emerges an ergodic distribution polarizing into twin peaks of middle income and very high income regions. Viewed in terms of population well-being, this evidence suggests that two simultaneous processes occur: a slow process of catching up of the poorest regions and a process of shifting away of very rich regions.

Our findings raise some issues that need to be addressed in a future research. First of all, using weighted per capita income levels in PPS as a proxy of living standards has some *caveat*, already mentioned in previous sections. Although this is a reasonable and practical assumption when we focus on economic well-being, the paper still ignores intra-region income disparities. Moreover, if one believes that standard of living is a multidimensional concept, it seems logical to follow a multivariate approach. The statistical tools presented in this paper can be generalised along this line of research.

References

Aitkin, M. and Wilson, G. T. (1980). 'Mixture models, outliers, and the EM algorithm', *Technometrics*, 22, 325–331.

Azariadis C. and Drazen, A. (1990). 'Threshold Externalities in Economic Development', *Quarterly Journal of Economics*, 105, 501–526.

Baldwin, R., Ottaviano, G. and Martin, P. (2001). 'Global Income Divergence, Trade and Industrialization: The Geography of Growth Take-Offs', *Journal of Economic Growth*, 6, 5–37.

Bartholomew, D.J. (1981). *Mathematical Methods in Social Science*, Chichester, John Wiley, New York.

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

Bianchi, M. (1997). 'Testing for convergence: evidence from non-parametric multimodality tests', *Journal of Applied Econometrics*, 12, 393–409.

Boldrin, M. and Canova, F. (2001). 'Inequality and convergence: reconsidering European regional policies', *Economic Policy*, 32, 205–53.

Bowman, A.W. (1992). 'Density based tests of goodness-of-fit', *Journal of Statistical Computation and Simulation*, 40, 1–13.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford University Press, Oxford.

Bulli, S. (2001). 'Distribution dynamics and cross-country convergence: a new approach', *Scottish Journal of Political Economy*, 48, 226–243.

Caniels, M.C.J., Verspagen, B. (2001). 'Barriers to knowledge spillovers and regional convergence in a evolutionary model', *Journal of Evolutionary Economics*, 11, 307–329.

Cheshire, P.C. and Carbonaro, G. (1995). 'Convergence-divergence in regional growth rates: an empty black box?', in: Vickerman R.W., Armstrong H.W. (eds), *Convergence and Divergence among European Regions*, Pion, London.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). 'Maximum likelihood from incomplete data via EM algorithm', *Journal of the Royal Statistical Society*, B, 69, 1–38.

Desgoits, A. (1999). 'Patterns of economic development and the formation of clubs', *Journal of Economic Growth*, 4, 305–330.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability*, 57, Chapman and Hall, New York.

Fan, Y. (1994). 'Testing the goodness of fit of a parametric density function by kernel method', *Econometric Theory*, 10, 316–356.

Fan, Y. (1995). 'Bootstrapping a consistent non parametric goodness-of-fit test', *Econometrics Reviews*, 14, 367–82.

Fan, Y. and Ullah, A. (1999). 'On Goodness-of-fit Tests for Weakly Dependent Processes Using kernel Method', *Journal of Nonparametric Statistics*, 11, 337–360.

Fingleton, B. (1997). 'Specification and testing of Markov chain models: an application to convergence in the European Union', *Oxford Bulletin of Economic and Statistics*, 59, 385–403.

Fingleton, B. (1999). 'Generalised linear models, loglinear models and regional dynamics', in: Dahiya S.B. (ed) *The Current State of Economic Science*, vol 1, Spellbound Publications, Rohtak, India.

Galor, O. (1996). 'Convergence? Inferences from theoretical models', *Economic Journal*, 106, 1056–69.

Howitt, P. (2000). 'Endogenous growth and cross-country income differences', *American Economic Review*, 90, 829–46.

Izenman, A.J. and Sommer, C.J. (1988). 'Philatelic mixtures and multimodal densities', *Journal of the American Statistical Association*, 83, 941–953.

Johnson, P.A. (2000). 'A nonparametric analysis of income convergence across the US states', *Economics Letters*, 69, 219–223.

Jones, C.I. (1997). 'On the Evolution of the World Income Distribution', *Journal of Economic Perspectives*, 11, 19–36.

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). 'A brief survey of bandwidth selection for density estimation', *Journal of the American Statistical Association*, 91, 401–407.

Kuznets, S. (1955). 'Economic growth and income inequality', *American Economic Review*, 45, 1–28.

Le Gallo, J. (2004). 'Space-time analysis of GDP disparities among European regions', *International Regional Science Review*, 27, 138–163.

Li, Q. (1996). 'Nonparametric testing of closeness between two unknown distribution functions', *Econometric Reviews*, 15, 261–274.

López-Bazo, E., Vayá, E., Mora, A.J. and Surinach J. (1999). 'Regional economic dynamics and convergence in the European Union', *Annals of Regional Science*, 33, 343–370.

Magrini, S. (1999). 'The evolution of income disparities among the regions of the European Union', *Regional Science and Urban Economics*, 29, 257–281.

Marron, J.S. and Schmitz, H.P. (1992). 'Simultaneous density estimation of several income distributions', *Econometric Theory*, 8, 476-488.

Marron, J.S. and Wand, M.P. (1992). 'Exact mean integrated squared error', *The Annals of Statistics*, 2, 712–736.

Martin, P. and G. Ottaviano (2001). 'Growth and Agglomeration', *International Economic Review*, 42, 947–968.

Maurseth, P.B. (2001). 'Convergence, Geography and Technology', *Structural Change and Economic Dynamics*, 12, 247–276.

McLachlan, G.J. (1987). 'On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture', *Journal of the Royal Statistical Society*, C, 36, 318–324.

McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Models*, New York, Wiley.

Paap, R. and van Dijk, H.K. (1998). 'Distribution and mobility of wealth of nations', *European Economic Review*, 42, 1269–1293.

Paci, R. (1997). 'More similar and less equal: economic growth in the European Regions', *Weltwirtschaftliches Archiv*, 133, 609–634.

Quah, D. (1993). 'Empirical cross-section dynamics in economic growth', *European Economic Review*, 37, 426–434.

Quah, D. (1996a). 'Empirics for economic growth and convergence', *European Economic Review*, 40, 1353–1375.

Quah, D. (1996b). 'Regional convergence clusters in Europe', *European Economic Review*, 35, 951–958.

Quah, D. (1996c). 'Convergence empirics across economies with (some) capital mobility', *Journal of Economics Growth*, 1, 95–124.

Quah, D. (1997). 'Empirics for growth and distribution: polarization, stratification, and convergence clubs', *Journal of Economics Growth*, 2, 25–59.

Quah, D. (2001). 'Searching for prosperity a comment', *Carnegie-Rochester Conference Series on Public Policy*, 55, 305–319.

Quinn, B.G., McLachlan, G.J., and Hjort, N.L. (1987). 'A note on the Aitkin-Rubin approach to hypothesis testing in mixture models', *Journal of the Royal Statistical Society*, B, 49, 311–314.

Reichlin, L. (1999). 'Discussion of Convergence as Distribution Dynamics', in: Baldwin R.D, Cohen D., Sapir A., Venables A. (eds), *Market Integration, Regionalism, and the Global Economy*, Cambridge University Press and CEPR, Cambridge.

Richardson, S. and Green, P.J. (1997). 'On Bayesian analysis of mixtures with unknown number of components', *Journal of Royal Statistical Society*, B, 731–792.

Sheather, S.J. and Jones, M.C. (1991). 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of Royal Statistical Society*, B, 53, 683–690.

Silvermann, B. W. (1981). 'Using Kernel Density Estimates to Investigate Multimodality', *Journal of Royal Statistical Society*, B, 43, 97–99.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Tsionas, E. G. (2000). 'Regional growth and convergence: evidence from the United States', *Regional Studies*, 34, 231–38.