STATISTICA, anno LXX, n. 1, 2010

INTRA-DISTRIBUTION DYNAMICS OF REGIONAL PER-CAPITA INCOME IN EUROPE: EVIDENCE FROM ALTERNATIVE CONDITIONAL DENSITY ESTIMATORS

Roberto Basile

1. INTRODUCTION

The interest in regional convergence has been growing intensively in the last decade. The most widely accepted method of testing the convergence hypothesis is the regression or β-convergence approach. This method has been discussed from different points of view (see Durlauf *et al.*, 2005, for a review of the literature on economic convergence; and Magrini, 2004, for a survey focusing on regional convergence studies). One of the critical points is that it focuses on the behavior of the representative economy. In particular, it sheds light on the transition of this economy towards its own steady state, but it does not provide any insight on the dynamics of the whole cross-sectional distribution of regional per-capita incomes. In fact, a negative relationship between the growth rates and the initial conditions can be associated with a rising, a declining and a stationary cross-section income dispersion. Clearly, a method that cannot differentiate between convergence, divergence and stationarity is of limited or no use. This failure is essentially a simple intuition of what is termed Galton's fallacy (Quah, 1993).

More recently, an alternative approach to the analysis of convergence has been proposed in order to solve this problem. Such a method, known as the *intra-distribution dynamics* approach (Quah, 1996a, 1996b, 1996c, 1997, 2007), examines directly how the whole income distribution changes over time and, thus, it appears more informative than the regression approach.

The intra-distribution dynamics was generally analyzed through the application of Markov chain methodologies (Quah, 1996b; López-Bazo *et al.*, 1999; Fingleton, 1997, 1999; Bulli, 2001) or, more recently, through the estimation of conditional densities using stochastic kernel estimators (Quah, 1997; Lamo, 2000; Pittau and Zelli, 2006; Magrini, 2004). All the studies using kernel estimators provide contour plots of the conditional density to describe the law of motion of cross-sectional distributions. So, they treat the conditional density function as a bivariate density function, while it has been noticed that the conditional density function is a sequence of univariate functions (Hyndman, 1996). Furthermore, these studies scantly take account of the recent developments in the statistical literature on conditional density estimation (Hyndman *et al.*, 1996; Fan *et al.*, 1996; Hall *et al.*, 1999; Hyndman and Yao, 2002), which have highlighted the strong bias problems associated with the widely used standard kernel estimator and have proposed new estimators with better statistical properties.

The aim of this paper is to explore alternative conditional density estimators and alternative graphical methods, both developed by Hyndman *et al.* (1996), to describe the law of motion of cross-regional distributions of per-capita incomes in Europe. In particular, Hyndman *et al.* (1996) have noticed that the mean function of the kernel density estimator is equivalent to the Nadaraya-Watson kernel smoother. Because of the undesirable bias properties of this smoother, they have proposed a modified conditional density estimator with a mean equivalent to some other nonparametric regression smoothers that have better statistical properties in terms of mean-bias. This new estimator has smaller integrated mean square error than the standard kernel estimator.

The analysis performed in the paper can be interpreted as a test of the hypothesis of "absolute convergence", since it does not control for the heterogeneity in the structural characteristics of the regions (for example, in terms of technologies, rates of population growth, saving rates and so on) (see Galor, 1996, for a distinction between absolute, conditional and club convergence hypotheses). Yet, assessing absolute convergence across European regions in terms of per-capita GDP is still a matter of primary importance in order to evaluate the effectiveness of Cohesion Policies. This hypothesis seems to be the one that the European Commission is interested in, as Quah (1996a, p. 1048, footnote 4) already pointed out.

The layout of the paper is the following. In Section 2, the most recent literature on the intra-distribution dynamics approach and on conditional density estimators is reviewed. In Section 3, estimation results obtained applying different estimators to EU regions' per-capita GDP data over the period 1980-2002 are reported. Section 4 concludes.

2. INTRA-DISTRIBUTION DYNAMICS AND DENSITY ESTIMATORS

2.1 *The transition dynamics approach and the kernel conditional density estimator*

Danny Quah (1997, 2007) has suggested an interesting approach to the analysis of economic convergence based on the concept of transition dynamics. In a nutshell, this method consists of studying the dynamics of the entire distribution of the levels of per-capita income of a set of economies. In the most recent version, it consists of estimating univariate and conditional density functions and of computing the ergodic distribution to describe the long-run growth behaviour of per-capita income distribution. The univariate analysis allows to identify the features of the regional distribution of per-capita income at different points in time (for example, at the initial and final years of a long time period). The conditional density analysis gives information on the changes of the relative position of various regions in the cross-section distribution of per-capita income over time, the so-called 'intra-distribution dynamics' (*IDD*).[1]

More formally, the *IDD* approach consists of estimating and visualizing the conditional density of Y given X, where Y is the regional per-capita income at time $t+\tau$ and X the regional per-capita income at time t. Denote the sample by $\{(X_1, Y_1),...,(X_n, Y_n)\}$ and the observations by $\{(x_1, y_1),...,(x_n, y_n)\}$; thus, the aim of the researcher is to estimate the density of Y conditional on $X = x$. Let $g_\tau(x, y)$ be the joint density of (X,Y), $h_\tau(x)$ the marginal density of X and $f_\tau(x \mid y) = g_\tau(x, y)/h_\tau(x)$ the conditional density of $Y \mid (X = x)$. The most obvious estimator of the conditional density is the kernel estimator, firstly proposed by Rosenblatt (1969). Recently, Hyndman et al. (1996) have further explored its properties. They define:

$$\hat{f}_\tau(x \mid y) = \hat{g}_\tau(x, y)\Big/\hat{h}_\tau(x) \tag{1}$$

where

$$\hat{g}_\tau(x, y) = \frac{1}{nab}\sum_{i=1}^{n} K\left(\frac{\left\|x - X_i\right\|_x}{a}\right)\left(\frac{\left\|y - Y_i\right\|_y}{b}\right)$$

is the estimated joint density of (X,Y) and

$$\hat{h}_\tau(x) = \frac{1}{na}\sum_{i=1}^{n} K\left(\frac{\left\|x - X_i\right\|_x}{a}\right)$$

is the estimated marginal density.[2]

The two parameters *a* and *b* control the smoothness between conditional densities in the *x* direction (the smoothing parameter for the regression) and the smoothness of each conditional density in the *y* direction, respectively. As usual, small bandwidths produce small bias and large variance whereas large bandwidths give large bias and small variance. As shown in Bashtannyk and Hyndman (2001), optimal bandwidths can be computed by generalized cross validation (GCV).

The bandwidth *a* can either be fixed or it can vary as a function of the focal point *x*. When the data are not homogenously distributed over all the sample space (that is when there are regions of sparse data), a variable (or nearest-neighbor) bandwidth is recommended. In this case, we adjust *a(X)* so that a fixed

[1] Within this approach, the intra-distribution dynamics is modeled as a stochastic kernel under the hypotheses of time invariance and first order evolution, that is the transition mechanism is assumed to be time invariant and the model is assumed to be a time-homogenous Markov Chain.

[2] $\|.\|_x$ and $\|.\|_y$ are Euclidean distance metrics on the spaces of *X* and *Y* respectively. *K(.)* is a symmetric density function, known as the kernel function. Usually, the Epanechnikof or the Gaussian kernel are used.

number of observations m is included in the window. The fraction m/n is called the span of the kernel smoother.[3]

Equation (1) can also be written as:

$$\hat{f}_\tau(y \mid x) = \frac{1}{b}\sum_{i=1}^{n} w_i(x) K\left(\frac{\left\| y - Y_i \right\|_y}{b}\right) \tag{2}$$

where

$$w_i(x) = K\left(\frac{\left\| x - X_i \right\|_x}{a}\right) \Bigg/ \sum_{j=1}^{n} K\left(\frac{\left\| x - X_j \right\|_x}{a}\right).$$

Equation (2) suggests that the conditional density estimate at $X = x$ can be obtained by summing the n kernel functions in the Y-space, weighted by $\{w_i(x)\}$ in the X-space. In other words, equation (2) can be interpreted as the Nadaraya-Watson kernel regression (or locally weighted averaging) of $K\left(\frac{\left\| y - Y_i \right\|_y}{b}\right)$ on X_i (see Hyndman and Yao, 2002). This estimator has two desirable properties: (*i*) it is always non-negative and (*ii*) integrals of the estimators with respect to y equal 1.

2.2 *A kernel conditional density estimator with mean-bias correction*

Hyndman *et al.* (1996) have observed that the estimation of the conditional mean function obtained from the kernel density estimator (Equation 2) is equivalent to the Nadaraya-Watson kernel regression function:

$$\hat{m}(x) = \int y \hat{f}_\tau(y \mid x) dy = \sum_{i=1}^{n} w_i(x) Y_i \tag{3}$$

As is well known, the Nadaraya-Watson smoother can present a large bias both on the boundary of the predictor space, due to the asymmetry of the kernel neighbourhood, and in its interior, if the true mean function has substantial curvature or if the design points are very irregularly spaced.

Given the undesirable bias properties of the kernel smoother, Hyndman *et al.* (1996) has proposed an alternative conditional density estimator with a mean function equivalent to that of other nonparametric regression smoothers having better properties than the Nadaraya-Watson one.

The new class of conditional density estimators can be defined as

[3] It is important to say that there is not any statistical method to choose the optimal span for the conditional kernel density estimation. A simple way for choosing the span, followed in this paper, is to start from the value of the span selected by *GCV* for the *k*-nearest-neighbor local linear smoother (see section 3.1) and then try with other values higher and lower than this value (see Section 3.3).

$$\hat{f}_\tau^*(y \mid x) = \frac{1}{b}\sum_{i=1}^{n} w_i(x) K\left(\frac{\left\| y - Y_i^*(x) \right\|_y}{b}\right) \tag{4}$$

where $Y_i^*(x) = e_i + \hat{r}(x) - \hat{l}(x)$, $\hat{r}(x)$ is an estimator of the conditional mean function $r(x) = E[Y \mid X = x]$, $e_i = Y_i - \hat{r}(x_i)$ and $\hat{l}(x)$ is the mean of $\hat{f}_\tau^*(e \mid x)$.

Since the error term (e_i) has the same distribution of y_i except for a shift in the conditional mean, one may start by applying the standard kernel density estimator to the points $\{x_i, e_i\}$ and, then, adding the values of $\hat{r}(x)$ to the estimated conditional densities $\hat{f}_\tau^*(e \mid x)$ in order to obtain an estimate of the conditional density of $Y \mid (X=x)$. Since $\hat{l}(x)$ is constant under certain conditions (homoskedastic and independent errors), the mean-bias of $\hat{f}_\tau^*(y \mid x)$ is simply the bias of $\hat{r}(x)$ and the integrated mean square error is reduced.

Obviously, setting $\hat{r}(x) = \hat{m}(x) = \sum_{i=1}^{n} w_i(x) Y_i$ (that is the Nadaraya-Watson smoother) implies that $\hat{f}_*(y \mid x) = \hat{f}(y \mid x)$. However, $r(x)$ can also be estimated by using many other smoothers having better properties than the kernel regression estimator, $\hat{m}(x)$.[4] In other words, using the method developed by Hyndman *et al.* (1996), the mean function of $\hat{f}_\tau^*(y \mid x)$ is allowed to be equal to a smoother with better bias properties than the kernel regression. In this way, we obtain an estimate of the conditional density with a mean-bias lower than that of the kernel estimator.

3. SOME EVIDENCE ON REGIONAL CONVERGENCE IN EUROPE

3.1 *Data, scatterplot smoothing and empirical strategy*

We analyze the IDD of regional per-capita incomes in Europe over the period 1980-2002. Per-capita income levels are normalized with respect to the EU average in order to remove co-movements due to the European wide business cycle and trends in the average values. The income variable is the total gross value-added (GVA) computed according to the European System of integrated Accounts (ESA95). The total GVA figures are at constant prices 1995 and are converted to Purchasing Power Standards (PPS). However, only national PPS have been applied, since Eurostat does not possess comparable regional price levels that would enable us to take regional differences in price levels into account. The number of NUTS2

[4] Using $\hat{r}(x)$ we often introduce an extra smoothing parameter, *c*. Notice that both *c* and *a* control smoothness in the *x* direction; *a* controls how quickly the conditional densities can change in shape and spread while *c* controls the smoothness of the mean of the conditional densities over *x*.

regions included in the sample is 189 (the list of regions can be provided upon request). Data are drawn from the Cambridge Econometrics Dataset.[5]

In order to estimate conditional density functions $f_\tau(y|x)$, evaluation at a large number of points is frequently required. For this reason, we fix $\tau = 15$ and exploit the panel structure of the dataset. Thus, Y and X are vectors of 1,512 observations (189 regions × 8 periods).[6]

Figure 1 shows the scatterplot of relative per-capita income levels at time t and $t+15$. We can clearly observe three things: (1) data are distributed around the main diagonal, indicating a high degree of immobility; (2) at the extreme of the sample space data are sparser; (3) a few extreme observations appear on the right side of the scatterplot. These six points refer to Groningen, a region often excluded from convergence analyses, since it always appears as an outlier.[7] However, in spirit of the distribution dynamics approach described by Quah (1997, p. 34), we did not exclude regions from the dataset just because they have "*performed extraordinary well or extraordinary poorly relative to the bulk of other macroeconomies*". They represent real people and real regions not just observations that might be useful to delete in statistical analysis. Rather, an effort has to be made to find estimation methods robust against outliers.

In Figure 1 we also superimpose the estimated fit of three different scatterplot smoothers: (a) the Nadaraya-Watson (kernel regression) estimator ('dotted' curve) with a Gaussian kernel and a fixed bandwidth h=0.097; (b) the local linear regression smoother ('long-dashed curve') with a variable bandwidth (also known as the k-nearest-neighbor local linear smoother)[8]; and (c) the *lowess* ('solid' curve)[9]. All

[5] In alternative to the NUTS regions, some authors have used Functional Urban Regions (FURs) as units of analysis (Magrini, 1999) in order to take the spatial sphere of socio-economic influence of any basic unit into account. However, the main data sources (Eurostat and Cambridge Econometrics) only provide data at NUTS level.

[6] In this kind of analysis, starting from the work of Quah (1997), it is standard to pool different cross-sections of the data to estimate conditional densities (see also Magrini, 2004). The choice of τ = 15 might appear as arbitrary. In Section 3.3 the results of the analyses with different values for τ are discussed.

[7] Groningen seems to have worsened its relative economic position in the second half of the eighties. However, the evolution of gas prices and changes in the way in which GDP in the energy sector was distributed between regions are well-known reasons for this feature. Thus, Groningen could not be considered as an economic outlier in strict sense and might be excluded from the analysis. However, in the present paper we decided to keep this region within the sample in order to show the potential effects of outliers on the estimate of conditional densities. In Section 3.3 we report the results of the analysis performed after having excluded Groningen from the sample.

[8] It is important to report that the result of the local linear regression smoother with fixed bandwidth is analogue to that of the kernel estimator with fixed bandwidth, while the result of the kernel regression smoother with variable bandwidth is analogue to that of the k-nearest-neighbor local linear smoother. Therefore, these two "intermediate" cases are not considered in Figure 1.

[9] The *lowess* can be interpreted as a tri-cube kernel scatterplot smoother, able to capture local fluctuations in the density function of the independent variable (Cleveland, 1979; Cleveland and Devlin, 1988). The combination of three features - nearest neighbours, smoothed weight function (the tricube kernel) and local expected value formed via locally weighted regressions - helps the *lowess* regression outperform many other scatterplot smoothers. In particular, a local linear smoother is, *per se*, not robust against outliers. Only, the *lowess* is very robust against 'far out' observations, since it down-weights large residuals.

bandwidth parameters have been selected by using the GCV method. In the cases (b) and (c) the span that defines the size of the neighborhood in terms of a proportion of the sample size is equal to 0.15 (the width of the smoothing windows always contain the 15% of the data).

As expected, the Nadaraya-Watson (or local averaging) smoother appears more sensitive than the other two smoothers to extreme observations (Groningen) and to the data sparseness at the boundary. Moreover, a difference between the local linear regression with variable bandwidth and the *lowess* emerges only at the extreme right side of the sample space, confirming that only the *lowess* is resistant against isolated points.

In the rest of this section we report the results of different conditional density estimators. First, we estimate $\hat{f}_{15}(y \mid x)$ using a kernel estimator with a constant bandwidth parameter *a* (equation 2). In this first step we compare two alternative graphical techniques for visualizing the conditional density estimators: the traditional perspective and contour plots, on the one side, and the new 'stacked' and '*HDR*' plots (described in section 3.2), on the other. Then, we estimate conditional densities using two alternative methods: (*i*) a kernel density estimator with variable bandwidth; (*ii*) a kernel density estimator with variable bandwidth and mean bias correction (equation 4).[10]

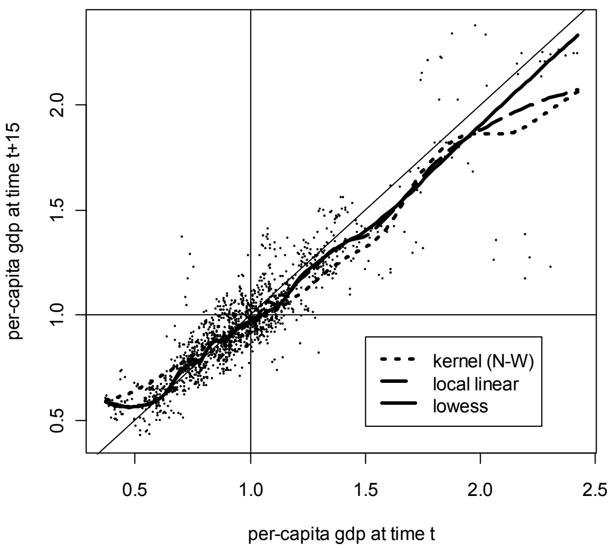


Figure 1 – Regional per-capita income in Europe: comparing different scatterplot smoothers.
Notes: the graph reports a scatterplot of relative per-capita income levels at time t and t+15. The estimated fits of three different scatterplot smoothers are superimposed: (a) the Nadaraya-Watson estimator ('dotted' curve); (b) the local linear regression smoother ('long-dashed curve') with variable bandwidth; and (c) the lowess ('solid' curve).

[10] All estimations were performed using the *R* software. In particular, we used the code *hdrcde* developed by Robert Hyndman and the code *locfit* (Loader, 1996). R scripts and data used in the paper are available upon request.

3.2 *New graphical methods for visualizing intra-distribution dynamics*

All the studies on intra-distribution dynamics using nonparametric stochastic kernel density estimators provide three-dimensional perspective plots and/or the corresponding contour plots of the conditional density to describe the law of motion of cross-sectional distributions. Therefore, they treat the conditional density as a bivariate density function, while the latter must be interpreted as a sequence of univariate densities of relative per-capita income levels conditional on certain initial levels.

Here we use new graphical methods for visualizing conditional density estimators developed by Hyndman *et al.* (1996) and Hyndman (1996). The first graphical technique, called the "*stacked conditional density plot*" (Figures 3A), displays a number of conditional densities plotted side by side in a perspective plot.[11] It facilitates viewing the changes in the shape of the distributions of the variable observed at time $t+\tau$ over the range of the same variable observed at time *t*. In other words, like a row of a transition matrix, each univariate density plot describes transitions over 15 years from a given income value in period *t*. Hyndman *et al.* (1996) note that this plot is "*much more informative than the traditional displays of three dimensional functions since it highlights the conditioning*" (p. 13).

The second type of plot proposed by Hyndman *et al.* (1996) is the "*highest conditional density region*" (*HDR*) plot (Figures 3B). Each vertical band represents the projection on the *xy* plan of the conditional density of *y* on *x*. In each band the 25% (the darker-shaded region), 50%, 75% and 90% (the lighter-shaded region) *HDR*s are reported. A high density region is the smallest region of the sample space containing a given probability. These regions allow a visual summary of the characteristics of a probability distribution function. In the case of unimodal distributions, the *HDR*s are exactly the usual probabilities around the mean value. In the case of multimodal distributions, the *HDR* displays different disjointed subregions.

The *HDR* plot is particularly important to analyze intra-distribution dynamics. If the 45-degree diagonal crosses the 25% or the 50% *HDR*s, it means that most of the elements in the distribution remain where they started (there is *'strong' persistence*); if it crosses only the 75% or the 90% *HDR*s, we can conclude in terms of *'weak' persistence*. If the horizontal line traced at the zero-value of the period *t+15* axis crosses *all* the 25-50% (75-90%) *HDR*s, we can say that there is *'strong' ('weak') global convergence* towards equality. Finally, if *some* 25-50% (75-90%) *HDR*s are crossed by a horizontal line traced at any value of the *t+15* axis, we can say that there is *'strong' ('weak') local or 'club convergence'*.[12] Clearly, this method is particularly informative for the analysis of regional growth behavior, since it highlights the dynamics of the entire cross-section distribution. It remains important to ana-

[11] Each univariate density plot is always non-negative and integrates to unity. Since the conditional density plot has been evaluated on an equispaced grid of 100 values over the range of *x* and *y* directions, Figure 3A displays 100 stacked univariate densities.

[12] The 'club convergence hypothesis' states that regions catch up with one another but only within particular subgroups.

lyze any other moment of the distribution (such as the mean and the variance) and any other central point. In particular, one may wish to analyze the modes, the values of y where the density function takes on its maximum values. In fact, especially when the distribution function is bimodal, the mean and the median are not very useful, since they will provide only a 'compromise' value between the two peaks. Thus, the modes may be considered as a form of robust nonparametric regression. In each figure, the highest modes for each conditional density estimate are superimposed on the *HDR* plots and shown as a bullet.

3.3 *Empirical evidence*

Figure 2 shows traditional perspective and contour plots for the conditional kernel density estimate with fixed bandwidth, describing 15-year horizon evolutions of the distribution of per-capita income relative to the European average. As well-known, the selection of the bandwidth parameter is a crucial issue in the estimation of densities. Optimal bandwidths have been selected using the method developed by Bashtannyk and Hyndman (2001) based on GCV. The bandwidth a for the x direction is 0.132, while the bandwidth b for the y direction is 0.127. This graph would suggest that over the period considered European regions have followed a convergence path. In fact, using the standard terminology, we observe a clockwise shift in mass indicating some degree of intra-distribution mobility, which would imply that the richer regions became poorer and the poorer became richer. These findings appear consistent with those reported in previous work (see, for example, Brasili and Gutierrez, 2004). Moreover, as it is common in these kinds of analyses, a 'multiple-peaks' property manifests. In fact, we can observe some distinct local maxima (or 'basins of attraction'). Contour plot makes this clearer.

The same estimation results discussed above are visualized in Figure 3 using the alternative stacked density plot and the *HDR* plot. From this graph, we would learn that regions that at the beginning of the period had a per-capita income level lower (higher) than the EU average would be more likely to improve (worsen) their relative position over the next 15 years: the 25% *HDR*s associated with relative per-capita income levels at time t lower (higher) than 1.0 (that is the European average) are all above (below) the main diagonal. Again, this means that the poorer economies would be catching up with the richer ones.

The *HDR* plot allows to identify (better than the standard contour plot) the presence of different 'convergence clubs'. The position of the highest modes and of the 25% *HDR*s would suggest local convergence at relative income levels of 0.7, 1.3, 1.8 and 2.2. Moreover, signs of bimodality would appear for very high levels of the distribution at time t: regions that at the beginning of the period had a very high income level would have experienced over time either a slowdown or a convergence towards the relative income of 2.2. However, this evidence is misleading, since a few extreme observations (all referring to Groningen) determines such an evidence of bimodality.

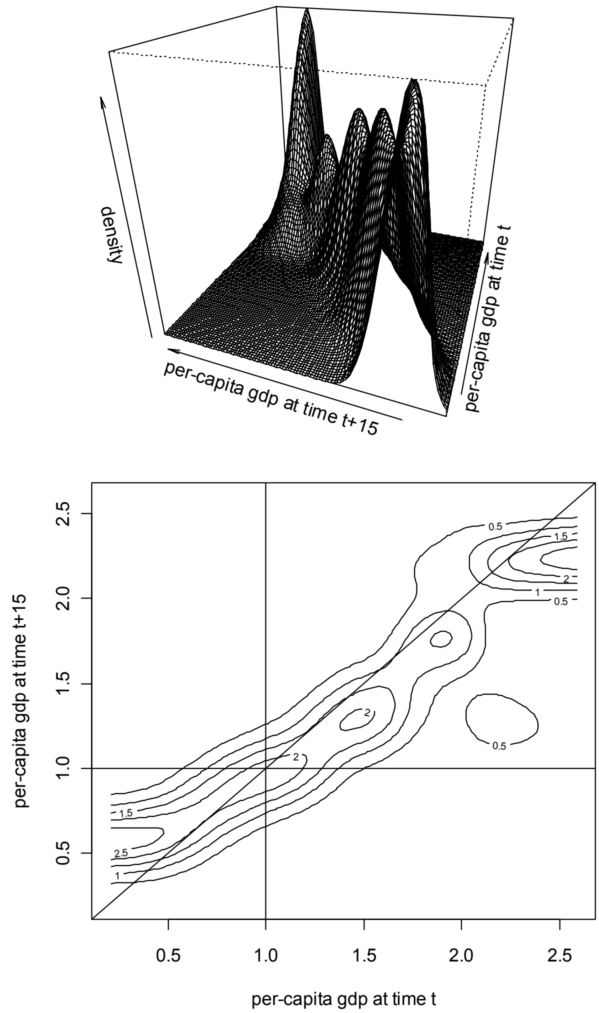
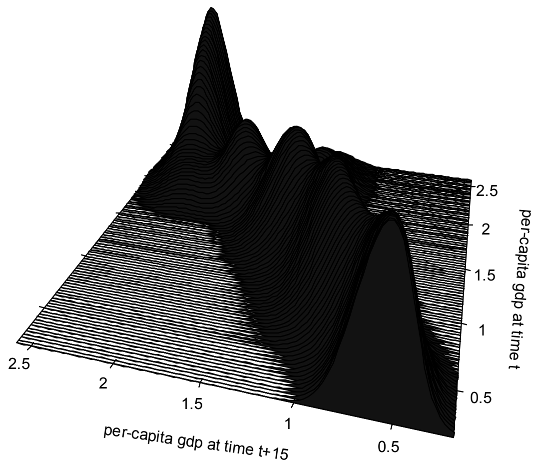


Figure 2 – Intra-Distribution Dynamics of regional per-capita income in Europe.
Standard perspective plot (left hand side panel) and contour plot (right hand side panel) of conditional density for transitions of 15 years between 1980-2002. Estimates are based on a kernel density estimator with fixed bandwidths (a = 0.132; b = 0.127).

Moreover, looking more carefully at Figure 3, we may observe that the plotted conditional density function does not fit the scattered points very well. In particular, we suspect that the sparseness of data at the boundaries and the presence of extreme points (Groningen) might have affected the entire estimated conditional density function, as well as they have affected the conditional mean function. Thus, alternative estimation methods are needed. First, we try with a kernel density estimator with a variable bandwidth to accommodate the problem of data sparseness (Figure 4).

A - Stacked density plot



B - HDR plot

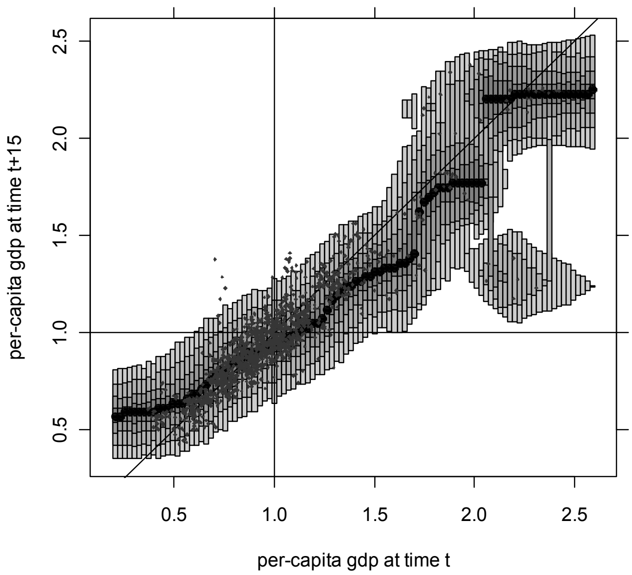


Figure 3 – Intra-Distribution Dynamics of regional per-capita income in Europe.
Stacked density plot and HDR plot of conditional density for transitions of 15 years between 1980-2002. Estimates are based on a kernel density estimator with fixed bandwidths (a = 0.132; b = 0.127).

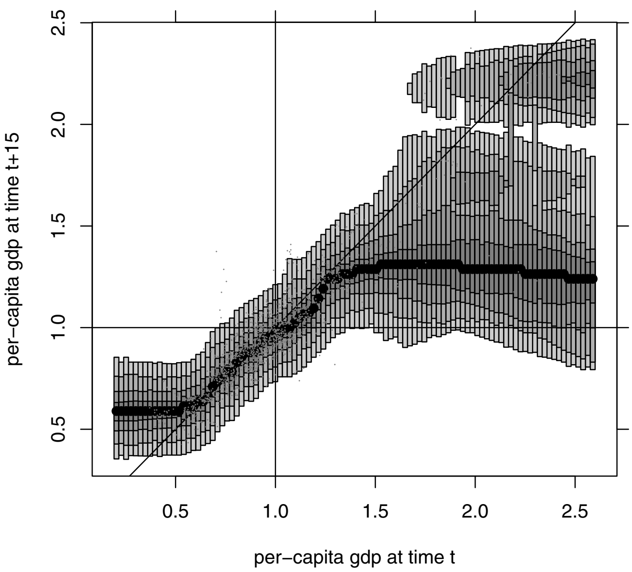


Figure 4 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 15 years between 1980-2002.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.15) and a fixed bandwidth in the y direction (b = 0.127).

The choice of a variable bandwidth substantially modifies the form of the conditional density function.[13] In particular, the evidence of mobility (and of convergence) is now confined to the upper and lower tails of the distribution at time *t*, while regions with a relative per-capita income between 0.7 and 1.3 at the beginning of the period did not change their relative position over time. However, the evidence of bimodality associated with very high initial income levels is now even stronger and, more importantly, the position of the highest modes is more strongly influenced by the values of Groningen. Therefore, the choice of a variable bandwidth seems to magnify the effect of outliers on the right hand side of the shape of the distribution. An estimator robust against outliers is definitely needed.

Thus, Figure 5 reports the results based on the modified conditional kernel density estimator with mean function specified by a *lowess* smoother. As it can be observed, after a certain threshold (about 0.6 times the European average), the 45-degree diagonal crosses the 25% and 50% *HDR*s and the modal regression follows a straight line. This reveals a high degree of immobility or persistence: European regions tended to maintain their relative positions over the study period. However, there is still some evidence of mobility at the left side of the sam-

[13] The span chosen for the variable bandwidth kernel density estimation is the same as that selected by GCV for the *k*-nearest-neighbours local linear smoother, that is 0.15. In section 5 we try with other span parameters to check the robustness of the results.

ple space: the 25% *HDR*s and the relative modes lie above the main diagonal for values of regional income lower than the threshold. This means that very poor regions registered higher growth rates than the other regions between 1980 and 2002. Moreover, these regions tend to converge towards a common level of relative per-capita income of about 0.6 times the overall mean, in line with the club convergence hypothesis. The convergence within this poorer group is shown by the slope of the modal regression which is almost parallel to the horizontal axis.

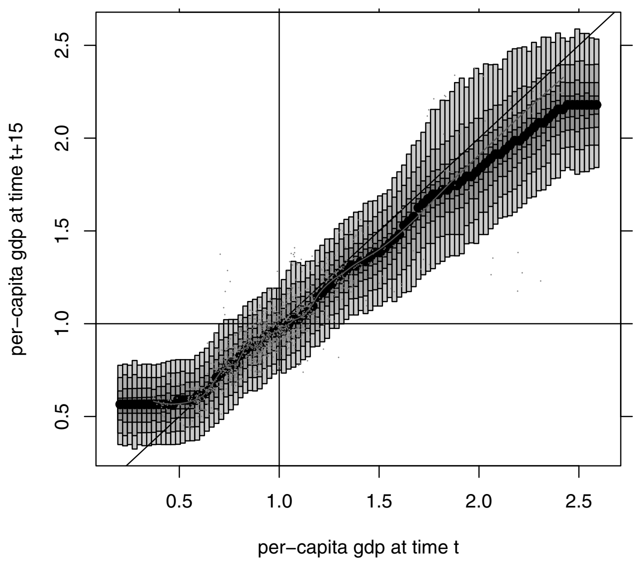


Figure 5 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 15 years between 1980-2002.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.15), a fixed bandwidth in the y direction (b = 0.127) and a mean function specified by a lowess smoother (span = 0.15).

3.3 *Robustness analysis*

In this section we briefly discuss the results of a number of further analyses aimed at checking the robustness of the findings discussed above. First of all, we discuss the results of an IDD analysis performed after having excluded Groningen from the sample. As already mentioned, the observed level of per capita GDP for this region is an artefact: it is the direct consequence of the way the Dutch accounting systems recorded revenues from oil extraction carried out in North Sea fields. Thus, both the exceptionally high level of per capita GDP at the beginning of the analysed period and the exceptionally low rate of growth over the 15-year periods have little or nothing to do with the way the regional economy actually worked. This means that the observation corresponding to Groningen cannot be interpreted, strictly speaking, as an outlier in economic sense. In contrast, it is just a measurement error, thus representing a typical example of observation that should be deleted in order to obtain sensible results. Since the aim of the paper is also to draw some effective conclusions on convergence dynamics of per capita income among European regions, here we check the robustness of the results reported in the previous section by replicating the last stage of the analysis (that is Figure 5) on a dataset that excludes Groningen. As in Figure 5, persistency and local convergence within a poor-regions' club seem to be the main messages of this estimate (see Figure A1 in the Appendix).

The second robustness check concerns the choice of τ. We have verified whether the results of the IDD analysis change as the transitional period changes from 15 years to 10 and 20. More specifically, we have considered τ =10 and τ = 20. Again, the main features depicted in Figure 5 remain unchanged (see Figures A2 and A3). The similarity between the short-horizon and the long-horizon intra-distribution dynamics can be interpreted as an evidence that the Markov property is somehow preserved (Quah, 2007).

The third issue that might raise some doubt on the reliability of the results reported in Figure 5 regards the choice of the adaptive bandwidth. In the case of Figure 5 we have used a variable *a* parameter for the estimation of conditional densities since the data are not homogenously distributed over all the sample space. That is we have adjusted *a(X)* so that a fixed number of observations is included in the window. As observed in Section 2.2, there is not a statistical method to choose the optimal span for the conditional kernel density estimation, yet. So, we have proceeded by using as span parameter for the conditional density the same value of the span selected by GCV for the *k*-nearest-neighbor local linear smoother (that is 0.15). Also this choice might appear as quite arbitrary. Thus, in order to check the robustness of the analysis, we have estimated the conditional density function with different values of the span: namely we tried with a span equal to 0.05 and 0.50. The results obtained (Figures A4 and A5) suggest that the IDD analysis of regional per capita GDP in Europe shown in Figure 5 is not influenced by the choice of the span parameter.

4. CONCLUSIONS

In order to describe the law of motion of cross-sectional distributions of regional per-capita incomes in the EU-15 area during the period 1980-2002 and following the intra-distribution dynamics approach proposed by Quah (1997, 2007), in this paper we have used a kernel density estimator with variable bandwidth and mean bias correction and the *Highest Density Regions* plot, developed by Hyndman *et al.* (1996) and Hyndman (1996). This density estimator is more robust against outliers and has better properties than the kernel estimator with fixed bandwidth traditionally used in the literature on intra-distribution dynamics. Moreover, the *Highest Density Regions* plot is very suitable for visualizing conditional density estimates, while the method generally applied in the literature (the contour plot) is more appropriate to display the joint distribution.

Applying the alternative methods to European regional data, we find evidence that enriches the debate on the distribution dynamics. In particular, we obtain evidence of strong persistency: over the period 1980-2002 most of the regions appear to remain where they were at the beginning; only a fraction of very poor regions improves its position over the time period converging towards a very low relative income level ('club convergence'). These results partially contrast with those reported in the literature on EU-15 regions (e.g., Brasili and Gutierrez, 2004) which indicate some degree of intra-distribution mobility, in the sense that poor regions become richer and rich regions grow less rapidly.

The analysis performed in this paper is essentially based on graphical inspection of the distribution. This may be considered as a weakness, since it would be useful to compare the conditional densities estimated with different methods through some indicators which help to sum up the degree of mobility of the distribution. This kind of indicators, such as the Shorrocks (1978) mobility Index have been used in the literature only for discrete analyses (transition matrices), but not within the continuous state-space approach. Even within the discrete framework, however, these indicators are not free of criticisms. The problem is that the results depend strongly on the number of classes used to recode the variables in a discrete space. Since the number of classes is generally arbitrarily chosen by the researcher, also the results of the mobility index turn out to be *ad hoc*. In the future, some effort will be devoted in order to overcome these problems and to better compare the estimated densities. Furthermore, some effort is needed to investigate the causes of persistence, using counterfactual analysis along the lines of Basile (2009). Also an extension of the analysis to the enlarged European Union is important.

ISAE (Institute for Studies and Economic Analyses) ROBERTO BASILE

APPENDIX

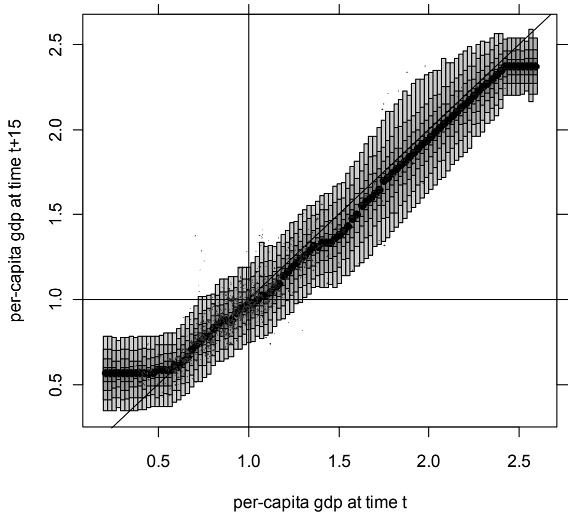


Figure A1 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 15 years between 1980-2002. Groningen is excluded from the sample.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.15), a fixed bandwidth in the y direction (b = 0.111) and a mean function specified by a lowess smoother (span = 0.15).

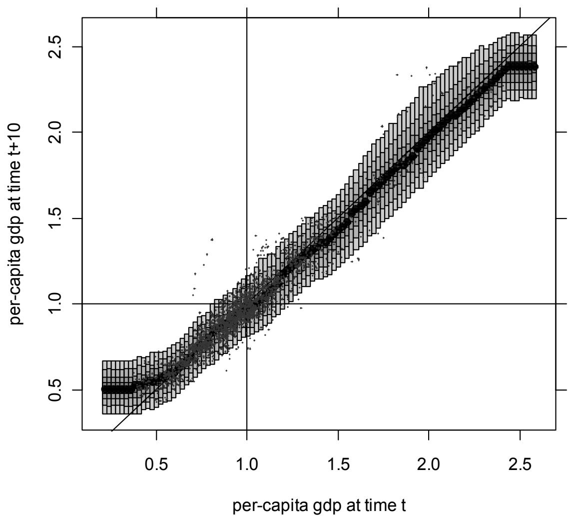


Figure A2 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 10 years between 1980-2002. Groningen is excluded from the sample.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.15), a fixed bandwidth in the y direction (b = 0.075) and a mean function specified by a lowess smoother (span = 0.15).

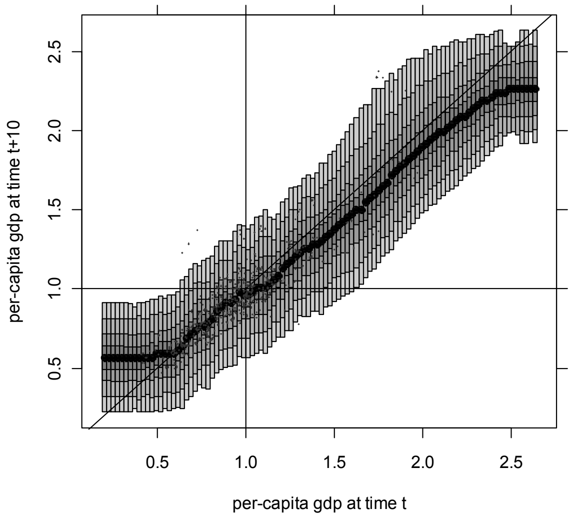


Figure A3 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 20 years between 1980-2002. Groningen is excluded from the sample.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.15), a fixed bandwidth in the y direction (b = 0.224) and a mean function specified by a lowess smoother (span = 0.15).

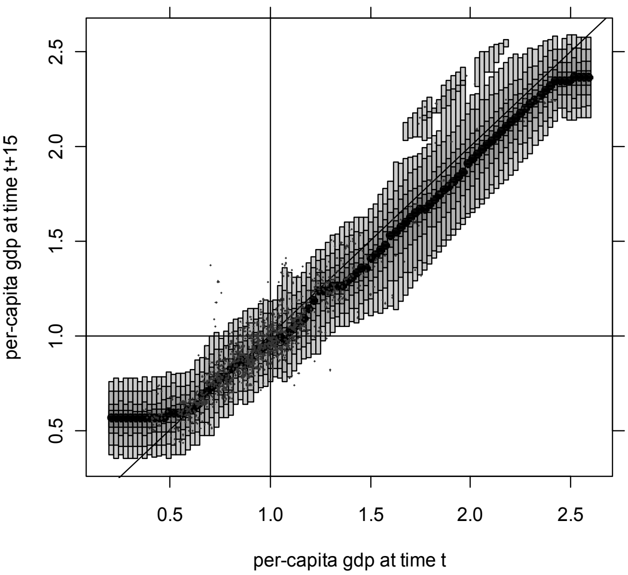


Figure A4 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 15 years between 1980-2002. Groningen is excluded from the sample.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.05), a fixed bandwidth in the y direction (b = 0.111) and a mean function specified by a lowess smoother (span = 0.15).

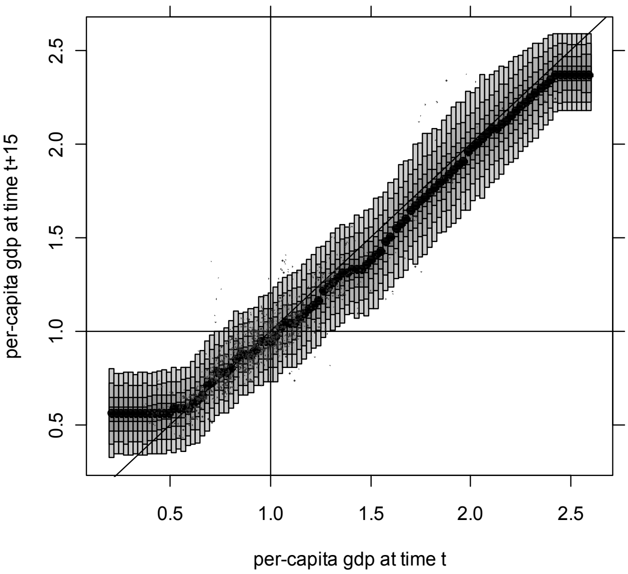


Figure A5 – Intra-Distribution Dynamics of regional per-capita income in Europe.
HDR plot of conditional density for transitions of 15 years between 1980-2002. Groningen is excluded from the sample.
Estimates are based on a kernel density estimator with a variable bandwidth in the x direction (span = 0.50), a fixed bandwidth in the y direction (b = 0.111) and a mean function specified by a lowess smoother (span = 0.15).

REFERENCES

R. BASILE (2009), *Productivity polarization across regions in Europe: The Role of Nonlinearities and Spatial*, "International Regional Science Review", 31, pp. 92-115.

D. BASHTANNYK, R. HYNDMAN (2001) *Bandwidth Selection for Kernel Conditional Density Estimation*, "Computational Statistics and Data Analysis", 36, pp. 279-98.

C. BRASILI, L. GUTIERREZ (2004), *Regional convergence across European Union*, "Development and Comp Systems 0402002, Economics Working Paper Archive".

S. BULLI (2001), *Distribution Dynamics and Cross-Country Convergence: New Evidence*, "Scottish Journal of Political Economy", 48, pp. 226-243.

W.S. CLEVELAND (1979), *Robust locally-weighted regression and scatterplot smoothing*, "Journal of the American Statistical Association", 74, pp. 829-836.

W.S. CLEVELAND, S.J. DEVLIN (1988), *Locally-weighted regression: an approach to regression analysis by local fitting*, "Journal of the American Statistical Association", 83, pp. 596-610.

S.N. DURLAUF, P.A. JOHNSON, J.R.W. TEMPLE (2005), *Growth Econometrics*, in P. Aghion, S.N. Durlauf (eds), "Handbook of Economic Growth ,Volume 1A", North-Holland, Amsterdam.

J. FAN, Q. YAO, H. TONG (1996), *Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems*, "Biometrika", 83, pp. 189-206.

B. FINGLETON (1997), *Specification and testing of Markov chain models: An application to convergence in the European Union*, "Oxford Bulletin of Economics and Statistics", 59, pp. 385-403.

B. FINGLETON (1999), *Estimates of Time to Convergence: An Analysis of Regions of the European Union*, "International Regional Science Review", 22, pp. 5-34.

O. GALOR (1996), *Convergence? Inferences from theoretical models*, "Economic Journal", 106, 1056-1069.

P. HALL, R. WOLFF, Q. YAO (1999), *Methods for estimating a conditional distribution function*, "Journal of American Statistical Association", 94, pp. 154-163.

R.J. HYNDMAN (1996), *Computing and Graphing Highest Density Regions*, "The American Statistician", 50, pp. 120-126.

R.J. HYNDMAN, D.M. BASHTANNYK, G.K. GRUNWALD (1996), Estimating *and visualizing conditional densities*, "Journal of Computational and Graphical Statistics", 5, 315-336.

R.J. HYNDMAN, Q. YAO (2002), *Nonparametric estimation and symmetry tests for conditional density functions*, 'Journal of Nonparametric Statistics", 14, pp. 259-278.

A. LAMO (2000), *On Convergence Empirics: Some Evidence for Spanish Regions*, "Investigaciones Economicas", 24, pp. 681-707.

C.R. LOADER (1996), *Local likelihood density estimation*, "The Annals of Statistics", 24, pp. 1602-1618.

E. LÓPEZ-BAZO, E. VAYÁ, A.J. MORA, J. SURIÑACH (1999), *Regional economic dynamics and convergence in the European Union*, "The Annals of Regional Science", 33, pp. 343-370.

S. MAGRINI (1999), *The Evolution of Income Disparities among the Regions of the European Union*, "Regional Science and Urban Economics", 29, pp. 257-281.

S. MAGRINI (2004), *Regional (Di)Convergence*, in V. HENDERSON, J.F. THISSE (eds.) "Handbook of Regional and Urban Economics" , North-Holland, Amsterdam.

M.G. PITTAU, R. ZELLI (2006), *Income dynamics across EU regions: empirical evidence from kernel estimator*, "Journal of Applied Econometrics", 21, pp. 605-628.

D. QUAH (1993), *Galton's fallacy and tests of the convergence hypothesis*, "Scandinavian Journal of Economics", 95, pp.427-443.

D. QUAH (1996a), *Twin Peaks: Growth and Convergence in Models of Distribution Dynamics*, "Economic Journal", 106, pp. 1045-55.

D. QUAH (1996b), *Empirics for economic growth and convergence*, "European Economic Review", 40, pp.1353–1375.

D. QUAH (1996c), *Convergence Emprics Across Economies with (Some) capital Mobility*, "Journal of Economic Growth", 1, pp. 95-124.

D. QUAH (1997), *Empirics for growth and distribution: stratification, polarization, and convergence clubs*, "Journal of Economic Growth", 2, pp. 27-59.

D. QUAH (2007), *Growth and distribution*, Mimeo LSE Economc department.

M. ROSENBLATT (1969), *Conditional Probability Density and Regression Estimators*, in P.R. KRISHNAIAH (ed.), "Multivariate Analysis II", Academic Press, NY.

S.J. SHEATHER, M.C. JONES (1991), *A reliable data-based bandwidth selection method for kernel density estimation*, "Journal of Royal Statistical Sociaty B", 53, pp. 683-690.

A. SHORROCKS (1978), *The measurement of mobility*, "Econometrica", 46, pp. 1013-1024.

SUMMARY

Intra-distribution dynamics of regional per-capita income in Europe: evidence from alternative conditional density estimators

In this paper different conditional density estimators are employed to analyze the cross-sectional distribution dynamics of regional per-capita income in Europe during the period 1980-2002. First, a kernel estimator with fixed bandwidth (the method traditionally

applied in the literature on intra-distribution dynamics) gives evidence of convergence. With a modified estimator, proposed by Hyndman et al. (1996), with variable bandwidth and mean-bias correction, the dominant income dynamics is that of persistence and lack of cohesion: only a fraction of very poor regions improves its position over time converging towards a low relative income ("poverty trap"). Moreover, an alternative graphical technique (more informative than the traditional contour plot) is applied to visualize conditional densities.