EXPLORING REGIONAL CONVERGENCE: EVIDENCE FROM 19 EUROPEAN COUNTRIES, 1991-2008
ANDRADE, Carlos*1
PINHO, Carlos*2,3
PINHO, Maria de Fátima2,4

Abstract
We focus on regional convergence based upon a sample of 19 European countries (1991-2008). After covering theoretical issues on $\beta$-convergence and $\sigma$-convergence, we focused on the income distribution details. Our evidence confirms some results presented elsewhere, such as the convergence process ephemeral nature, the trend towards polarization, the relative region's position high degree of persistence and the 'convergence clubs' hypothesis admissibility. Our main novelty is the estimated shape for the ergodic distribution, revealing a long run trend towards the formation of more clusters. In sum, the catching-up process underway in Europe seems to be transient.

JEL Classification: C14, O17, O47, O57, R11

Key words: regional growth; $\beta$-convergence; $\sigma$-convergence; distribution dynamics; Europe

1. - Introduction
The establishment of the European Union in 1957 put the reduction of regional income disparities as one of the Union's public policy explicit goals. Since then this focus has been successively reaffirmed along the deepening of the economic integration process. The enlargement underway along the last decades maintained the issue of European regional convergence under close focus. The overall aim has been the design of proper public policy tools to reduce the differences between the levels of development of the various European regions.

Closing attention upon the literature, the empirical evidence gathered worldwide has uncovered a set of stylized facts. Durlauf et al. (2005) mentioned the following three classes:

- Most countries have grown richer, but there are strong income disparities and substantial heterogeneity across countries;
- The international distribution of GDP per worker exhibits an emerging bimodality, which denotes little churning across the emerging twin modes;
- There appears to exist a plethora of factors that might affect growth, beyond factor accumulation and exogenous technical change which drove the Solow-Swan model.

According to Durlauf et al. (2008), the above mentioned stylized facts have correspondence in the following major streams of research:

---

*1 Instituto Superior de Contabilidade e Administração, Universidade de Aveiro, Aveiro. Portugal.
E-mail: cftandrade@ua.pt (corresponding author).2 Unidade de investigação em Governança, Competitividade e Políticas Públicas, Universidade de Aveiro, Aveiro. Portugal.3 Departamento de Economia, Gestão e Engenharia Industrial, Universidade de Aveiro, Aveiro. Portugal.E-mail: cpinho@ua.pt.4 Instituto Superior de Contabilidade e Administração, Universidade de Aveiro, Aveiro. Portugal.E-mail: fatima.pinho@ua.pt
1. One is focused on the convergence or divergence process. The main interest lies in the evaluation of the persistence of the differences between economies over long time horizons: are these differences transient or permanent? In case of permanent differences, a subsequent question is to assess if those differences reflect structural heterogeneity or initial conditions.

2. The second covers the properties of the cross-section. The focus is on the features of the probability density to describe incomes, namely on the evolution of the densities.

3. The third encompasses the identification of growth determinants. The attention is devoted to the explanation of the differences in growth, namely the factors that might be growth enhancers or determinants. Some issues are of special interest, such as the theoretical and empirical foundations for the determinants as well as a distinction between those factors that are fundamental and those that only are proximate.

The literature on European regional economic growth has fallen on those strands. Some authors devote their attention to the convergence vs. divergence discussion (Battisti and Di Vaio 2008; Debarsy and Ertur 2006; Fischer and Stirböck 2004; Lopez-Rodriguez 2008; Marelli 2007; Meliciani and Peracchi 2006), others are focused at intra-distributional dynamics (Basile 2009; Bosker 2009; Ezcurra and Rapun 2006; 2007; Fischer and Stumpner 2008; Hierro and Maza 2009; Le Gallo 2004; Maza and Villaverde 2004; Pittau and Zelli 2006; Villaverde and Maza 2008), while others are focalized at explaining the differences in growth (Becker et al. 2008; Dall'erba and Le Gallo 2008; Esposti and Bussoletti 2008; Ezcurra 2007; 2009; Guisan and Aguayo 2004; Le Gallo and Dall'erba 2008; Lopez-Rodriguez and Faina 2007; Maza et al. 2009; Mora 2008; Mulas-Granados and Sanz 2008; Olejnik 2008; Rodríguez-Pose and Crescenzi 2008; Zagler 2003).

This paper will develop along the second highlighted line of research. We aim to understand growth facts in the context of growth theories. Along this line of reasoning, we will undertake an exploratory analysis of convergence, focusing β-convergence and σ-convergence, as well as examining the evolution of the cross-sectional distribution of per capita income, through stochastic kernels. After presenting the different notions of convergence available, alongside with some aspects of growth econometrics and empirics, the empirical study will be presented aiming to uncover the regional convergence features of 19 European countries, using as regional units the European NUTS 2 definition. The methodology for the distribution dynamics approach draws heavily on Magrini (2007) and the empirical study is supported upon data for 1991-2008, retrieved from Cambridge Econometrics.

The results point to an overall process of regional convergence within the covered regions, following Marelli (2007). If this result is in line with the literature, our evidence on income per capita discloses a long run tendency to cluster across a wide set of values. Our evidence does not confirm the bimodal ergodic distribution detected by Basile (2009), while studying a sample of 15 European countries, or the unimodal ergodic distribution pointed out by Ezcurra and Rapun (2007), based on a sample of 25 European countries. Nonetheless, in the long run the results we present are consistent with the existence of the so-called ‘convergence clubs’, even if the set of clubs is somehow blurred.

The paper is organized as follows. Section 2 presents key theoretical aspects on
Andrade, C., Pinho, C., Pinho, M.F. *Exploring Regional Convergence in 19 European Countries*

convergence concepts. In Section 3, an account of the methodology and data used is provided, as well as a presentation and discussion of the obtained results. In the final section some conclusions are drawn.

2. Conceptualizing the convergence process

In the last half of the past century, the theoretical literature on economic growth has gone through several phases. Between the late 50s and the mid 80s, the literature was dominated by the Solow-Swan exogenous growth model, usually labeled as the neoclassical growth model. In the mid 80s, after Romer's (1986) contribution, the literature on economic growth refocused its attention with the introduction of the so-called endogenous growth theory and the subsequent discussion about the validity of the alternative theories. A central issue under discussion was whether income levels of poorer countries or regions were converging to those of richer ones. Hence, the undertaken empirical work on economic growth targeted the relative roles of factor accumulation and technical change in explaining growth trends.

Within the neoclassical growth framework, the diminishing marginal returns of capital was a crucial hypothesis to explain why richer economies would grow more slowly than poorer ones, where capital accumulation would still be under way. Based upon that assumption, the long run growth rate would rely only on the rate of technological change which is exogenous, ceteris paribus - i.e. assuming a constant population. Hence, the growth rate would be independent of any economic behavior and this would imply that economic policies would only produce temporary effects upon economic activity Cuaresma et al. (2008). While assuming that all countries or regions share in the technology progress equally, this approach established a negative relationship between the growth rate of per capita income and the initial level of per capita income.

The endogenous growth approach, relying on constant or increasing marginal returns to the factors as a fundamental hypothesis, assigns a central role to capital formation - empirical evidence points to a strong association in the data between investment and growth. Within this framework the concept of capital is no longer confined to the physical capital as in the neoclassical model, but it encompasses intangible capital, thus including human skills and the stock of knowledge. This wider concept plays an important role in the growth process because it reflects the response of economic actors in a competitive environment. Hence, technology became endogenous and subject to decision-making processes of individual agents. A set of endogenous factors - such as innovation, scale economies or learning processes - may produce externalities that spillover to the economy. A distinct role for government policies comes out with this approach, since public intervention should be designed to increase the accumulation of those factors that are responsible for generating externalities. Within this framework the convergence of per capita income between economies is no longer an inevitable outcome. There is a possibility for a country or region to go through a divergence process, in which the economy is "stuck" in a poverty trap, with low GNP per capita and low growth.

Islam (2003) noted that the debate over neoclassical growth theory and endogenous growth theory led to many different interpretations of convergence. We will start by focusing the concepts of β-convergence and σ-convergence, which represent one of many dichotomies that arose in the literature - see Islam (2003). After the discussion of these notions, we will close attention on the income distribution details, following Quah's
The convergence hypothesis is based upon the neoclassical growth models of Solow (1956) and Swan (1956). Drawing upon the presentation by Barro and Sala-i-Martin (1999), the main features of the neoclassical growth are summarized in the Annex.

The term $\beta$-convergence denotes the negative correlation between the growth rate of per-capita income over a given period of time and the initial per-capita income level. If the convergence is conceptualized in this way, the sign of the initial income variable in the so-called growth-initial level regression captures the negative correlation just mentioned. That regression is derived from the transitional dynamics of the model with Cobb-Douglas technology, through a Taylor series approximation around the deterministic steady state, which yields:

$$
\log y_t = \left[ \log y_0 - \left( \log y^* + \log A_0 \right) \right] e^{-\beta t} + \left( \log y^* + \log A_0 + \mu \right)
$$

where $y_t$ denotes the level of output per capita at an initial time, $y^*$ stands for the level of output per unit of effective labor that will be reached in the steady state and $\mu$ represents an error term that accounts for unpredictable effects on the growth rate. The average growth rate of per capita income over the period $T$ (where $T \geq 0$) starting at moment 0, is:

$$
\frac{1}{T} \log \left[ \frac{y_t}{y_0} \right] = \frac{1}{T} \left( 1 - e^{-\beta t} \right) \left( \log y^* + \log A_0 \right) + \frac{1}{T} \left( 1 - e^{-\beta t} \right) \log A_0 + \mu
$$

Hence, ceteris paribus, the average growth rate of per capita income is negatively related with its initial level, and it is conditioned on the exogenous growth rate of technology, on the steady state value of income per effective worker, and on the initial level of technology. Since the conditioning factors have an unknown form, the so-called growth-initial level regression usually has the following specification proposed by Barro and Sala-i-Martin (1991, 1992):

$$
\frac{1}{T} \log \left[ \frac{y_t}{y_0} \right] = \alpha_0 + \lambda_0 \log y_t + \mu
$$

where $y_{it}$ stands for per capita income of economy $i$ ($i=1,...,N$) at the moment $t$, $t$ is the time length and $\mu_i$ represents an error term that accounts for unpredictable effects on the growth rate.

This approach requires an appropriate set of variables to be included in the previous equation 3 in order to control the differences - see equation 8.

$$
\frac{1}{T} \log \left[ \frac{y_t}{y_0} \right] = \alpha_0 + \lambda_0 \log y_t + \psi X + \mu
$$

where $X$ stands for the set of explanatory variables that are the steady state determinants. For instance, Karras (2008) presents empirical findings that clearly reject absolute convergence in income per capita but are very strongly supportive of conditional
convergence. One must notice that $X$ represents those growth determinants that lie outside Solow and Swan's approach (Durlauf et al. 2005).

The $\beta$-convergence concept encompasses another relevant dimension for the understanding of convergence. We already covered the notions of unconditional convergence and conditional convergence. The first one implies one equilibrium level to which all economies close in. The latter involves different equilibria for each economy, since each one approaches its own and unique equilibrium. Hence, there might be space in this second view to conceive the possibility of a group of countries approaching a particular equilibrium if they share initial attributes that correspond to that equilibrium. In this case, the sense for the term convergence is that of club (or local) convergence and the model should yield multiple equilibria.

2.2. $\sigma$-convergence

Several methods have been employed to the distribution’s analysis. Initially, the research focused on how the dispersion of income differences developed over time. This notion of economic change, known as $\sigma$-convergence, occurs if the cross-sectional standard deviation of per capita income ($\sigma$) is falling over time. Hence:

$$\sigma_{t+1} - \sigma_t < 0$$

This notion continues to deserve attention until now, as it is shown by Villaverde (2006) and Young et al. (2008). Villaverde (2006) focuses on the Spanish regions to conclude that the high level of persistence justifies the need for a regional policy. The latter is centered on the so-called Gallon’s fallacy, which states that $\beta$-convergence implies $\sigma$-convergence. The work of Young et al. (2008) adds new evidence to this discussion, based upon a very large dataset. The fallacy has been proved erroneous, however, as these authors note once again; $\beta$-convergence is a necessary but not sufficient condition for $\sigma$-convergence.

Unlike $\beta$-convergence, statistical tests for $\sigma$-convergence were not readily available. An avenue of research has been focused on identifying regression specifications to build inferences for $\sigma$-convergence. However, great difficulties can be pointed out in interpreting tests of $\sigma$-convergence, since they assume that the data generating process is not invariant. Since evidence points to evolving distributions for the data, there is uneasiness about test distributions under the null hypothesis of $\sigma$-convergence.

Another issue that has deserved some recent attention is the formulation of a sensible notion of conditional $\sigma$-convergence. Durlauf et al. (2005) calls attention for the need of such a contribution. Pfaffermayr (2009) defines conditional $\sigma$-convergence as the average variance reduction of the predicted per capita income and introduces Wald tests for the conditional $\sigma$-convergence while using a maximum likelihood approach.

2.3. Income distribution dynamics

One way to study the features of income distributions is by estimating income densities with stochastic kernels and by analyzing their evolution over time. Rey and Janikas (2005) refer the use of stochastic kernels to study changes in modality, distributional mixing and stochastic dominance. Recently, Maza et al. (2009) also apply stochastic kernels, although with a variant that was suggested by Hyndman, Bashtannyk
and Grunwald, in 1996. This new approach is based on the estimation of the stacked density plot and highest density plot and aims to find variables that can partially explain differences in productivity whose nature was previously found persistent. One of the main results of the paper is that space matters, as well as investment per employee and compensation per employee.

Another path for the study of the income distributions features is through Markov chain modeling. Carrington (2006) used Markov chain as a starting point and departed from the literature on convergence by applying second-order stochastic dominance (SOSD) analyses. Carrington claims that SOSD overcomes the disadvantages of Markov chain modeling, namely the distortions brought by the discretization. Moreover, according to Carrington, SOSD allows more links between economy theory and statistical measurement, although these relations are better suited for utility and choice problems than growth issues.

Rey and Janikas (2005) mention a final approach to distributional modeling, which is based on regression trees. Labeled as a less ad hoc methodology, this approach allows the identification of convergence clubs as well as parameter heterogeneity in the underlying growth models.

A common critic pointed out to all the strategies for exploratory analyses of convergence is their disregarding for the implications of space. Spatial dependence has just begun to attract attention, as it is shown by Maza et al. (2009).

3. Empirical study

This paper’s approach to the convergence analysis is essentially exploratory, via the examination of the evolution of the cross-sectional distribution of per capita income, using stochastic kernels to describe both the change in its external shape and the intra-distribution dynamics. A detailed presentation of the distribution dynamics approach to be used will be delivered as well as information on the dataset. The results are presented afterwards.

3.1. Methodology

The approaches based upon the growth-initial level regressions have been thoroughly scrutinized by many scholars and their limitations have been emphasized. Magrini (2007) stresses two criticisms already pointed out: the models and the techniques discussed within those approaches produce results which are consistent with different explanations for growth, hence, the evidence gathered is of limited use due to its inability to discriminate between theories; the regression approach seems to produce little informative content, since it only describes how an economy converge to its own steady state, and nothing else. Moreover, this approach could be misleading, since the difference between conditional and unconditional convergence is not totally transparent, as some authors argue. These arguments support our option for this methodology.

The distribution dynamics approach implies the estimation of the operator $M_{t,s}$ highlighted above, from cross-sectional data on (relative) per capita income. The following presentation draws heavily on Magrini (2007).

If $x$ and $x'$ indicate generic values of respectively $x(t)$ and $x(t + s)$, then $p_{t,s}(x, x')$ is a non-negative measurable function that integrates to 1. When the stochastic kernel
$M_{t,s}(x,A)$ in equation 11 can be defined as the integral of $p_{t,s}(x,x')$ over the set $A$:

$$M_{t,s}(x,A) = \int_{-\infty}^{+\infty} p_{t,s}(x,x')dx'$$

(14)

the transition density function associated with the stochastic kernel is $p_{t,s}(x,x')$. Hence, the dynamics in equation 11 becomes:

$$f_{x(t+s)}(x') = \int_{-\infty}^{+\infty} p_{t,s}(x,x')f_{x(t)}(x)dx$$

(15)

Solving the equation 15 for $p_{t,s}(x,x')$, yields:

$$p_{t,s}(x,x') = \begin{cases} f_{x(t),x(t+s)}(x,x') \times f_{x(t)}(x)^{-1}, & \text{when } f_{x(t+s)}(x) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

(16)

and, therefore it is the probability density function of $x_{i}(t+s)$ conditional on $x_{i}(t)$. An estimate of the stochastic kernel can be obtained by dividing the estimate of the joint probability density function $\hat{f}_{x(t),x(t+s)}$ by the estimate of the marginal probability density function $\hat{f}_{x(t)}$:

$$\hat{p}_{t,s} = \frac{\hat{f}_{x(t),x(t+s)}}{\hat{f}_{x(t)}}$$

(17)

As pointed out before, there is a simple technique to proceed with the estimation, which is through the discretization of the income distribution (for instance, via the histogram). However, this technique has been ruled out due to its statistical inefficiency, especially if compared with other nonparametric estimators. As noticed before, such a discretization scheme can arbitrarily alter the probabilistic properties of the data.

In face of such critical remarks, the option would be to leave aside the discretization methodology and opt to estimate the stochastic kernels through kernel density estimators. Within this view, the stochastic kernel is interpreted as a transitional matrix with a continuum of rows and columns and the densities in equation 17. The analysis of convergence will be based on the shape of a three-dimensional plot:

- The main diagonal highlights persistence properties, hence the concentration along this line signals those elements of the cross-sectional distribution that remained where they started;
- The concentration around the 1-value of the time $t+s$ axis and parallel to the time $t$ axis signals the tendency towards convergence;
- The presence of separate modes (one, two or more) signals the tendency towards polarization.

The technical procedures on kernel density estimators and on the estimating procedure follow Magrini (2007). The bivariate kernel estimator employed to estimate the joint probability density function of $x(t)$ and $x(t+s)$ becomes:
\[
\hat{f}_x = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{h_{x_i(t)}} K \left( \frac{x(t) - x_i(t)}{h_{x_i(t)}} \right) \frac{1}{h_{x_i(t+s)}} K \left( \frac{x(t+s) - x_i(t+s)}{h_{x_i(t+s)}} \right) \right\}
\]

where \( h_{x_i(t)} \) is the bandwidth and \( K \) the kernel function (which is the bivariate Gaussian kernel function). One must note that the choice of the bandwidth matrix is quite crucial and there are still academic efforts to optimize its value.

### 3.2. Data

The dataset, retrieved from Cambridge Econometrics, displays the regional Gross Value Added (GVA) per capita as the income variable and spans the period 1980-2008 for the 252 NUTS\(^1\) from 19 different countries\(^2\). This sample figures are in millions of Euro constant prices, constant Purchasing Power Parities (PPP), for base year 2000.

The selection of countries in the sample was linked to data availability, since it was necessary to balance a wider regional coverage with a longer time frame track. Nonetheless, the countries selected may provide interesting insights over possible dichotomies on the performance paths of European countries, such as North-South or East-West dichotomies.

As for the source of the data, Cambridge Econometrics provides datasets which are based mainly on information supplied by REGIO, the Eurostat regional database. Unfortunately, REGIO's data displays serious handicaps in some respects. Cambridge Econometrics, through data retrieved from national statistics institutes and with the use of interpolation methods, was able to complete REGIO database. Hence, the popularity of this dataset in the literature - it offers a complete series of regional data for a sufficient number of NUTS2 regions over time.

Basile (2009), Bosker (2009), Ezcurra and Rapun (2006; 2007), Ezcurra et al. (2007), Fotopoulos (2008), Hierro and Maza (2009), Maza et al. (2009) and Villaverde and Maza (2008) also conducted empirical studies based on datasets provided by Cambridge Econometrics. To our knowledge, except for Ezcurra and Rapun (2007), those studies used samples which comprised 15 or less European countries. The only exception focused 25 countries.

### 3.3. Results

Our analysis focused initially on \( \beta \)-convergence. The results for the period 1991-2008, displayed on table 1, are in line with those presented in the literature, according to the survey by Dobson et al (2006). The estimation results signals the presence of a process of unconditional convergence across the European regions under study, with an estimated value of the rate of convergence towards the common steady state of 1.01\% and with the correspondent half-life of about 68 time periods.

---

\(^1\) Usually data is based on the Nomenclature of the Territorial Units for Statistics (NUTS) established by Eurostat, namely the NUTS-2 level, which stands as the common level of classification for the empirical analysis seen in the literature. However, some authors claim that, for some countries, NUTS-2 may not be adequate as an administrative level.

\(^2\) The set of countries is: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Romania, Sweden and United Kingdom.
Table 2 reports the standard deviation and the variation coefficient of the two series, in order to analyze what happened to $\sigma$-convergence. The table reported two test statistics used by Magrini (2007), which were proposed by Carree and Klomp. Both statistics show that the variance has decreased. Since both test statistics display low p-values, our confidence on the assertion of decreasing variance is strong. Hence, there is unambiguous evidence in favor of both unconditional $\beta$-convergence and unconditional $\sigma$-convergence. These results are in line with those by Marelli (2007), who did report a reduction of disparities (for both sigma and beta convergence).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-statistic</th>
<th>t-probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.107822</td>
<td>12.643066</td>
<td>0.000000</td>
</tr>
<tr>
<td>initial</td>
<td>-0.158242</td>
<td>-10.485122</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Ordinary least-squares estimate of beta

- $\beta = 0.0101$
- half-life = 68.4041

Non-linear least-squares estimate of beta

- $\beta = 0.0101$
- half-life = 68.4041

Table 2: $\sigma$-convergence over the period 1991-2008

<table>
<thead>
<tr>
<th></th>
<th>initial year</th>
<th>final year</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient of variation</td>
<td>0.0748</td>
<td>0.0634</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.7167</td>
<td>0.6271</td>
</tr>
<tr>
<td>mean</td>
<td>9.58</td>
<td>9.90</td>
</tr>
<tr>
<td>TEST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>value</td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td>p-level</td>
<td>53.96</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>0.000000</td>
<td>0.000003</td>
</tr>
</tbody>
</table>

A remark must be made over these results. According to Le Gallo (2004), Battisti and Di Vaio (2008), Fotopoulos (2008), Frenken and Hoekman (2006) and Maza et al. (2009) the geography has an important role in the convergence process. Moreover, as pointed out by Dobson et al (2006), one should use spatial econometric methods when cross-national studies are run. Without a proper assessment of space, the confidence over the above conclusion on convergence diminishes. Despite this, one must bear in mind that Frenken and Hoekman (2006), based on their empirical evidence, pointed out that usual
convergence conclusions old when a spatial econometric perspective is adopted. Also, Le Gallo and Dall'erba (2008) state that, while undertaking a σ-convergence, the conventional approach and the spatial one do not produce contradictory conclusions.

The estimates for the different density functions for 1991 and 2008 are shown in figure 1. In 1991 two modes could be identified. As in Ezcurra and Rapun (2007), besides the main mode, a second mode can be identified with a lower level of income. By 2008 those modes still persist, even though with less mass around the lower mode and with the first mode with slightly more mass and a little far away from the lower mode (figure A in appendix also displays a similar pattern). This may be an indication that the regions at the bottom of the distribution are converging towards a lower GVA per worker level than the rest of the sample, which would be consistent with the existence of the so-called `convergence clubs'. Hence, for now, our conclusion is quite similar with that taken by Ezcurra and Rapun (2007), as well as the one by Fischer and Stumpner (2008). Fischer and Stumpner focused on a sample of 27 European countries with data retrieved from Eurostat for 1995-2003. We must stress that those authors used a different kernel estimator from ours.

Now we will focus on the intra-distribution mobility during the study period, in order to address the issue of mobility of relative positions over time and assess the validity of the first impression retrieve from the analyses of β and σ-convergence, which was the tendency to decreasing disparities between European regions. Figure 2 displays the non-parametric estimate of the stochastic kernel for the transition dynamics across European regions throughout the study period. Figure 3 shows the contour plot of EUR-19 relative regional income distribution, along which the lines connect points at the same height on
the three dimensional kernel of figure 2.

As shown in figure 2 and especially on figure 3:

- a large portion of the probability mass is concentrated on the diagonal (the width of the iso-probs is very tiny for higher probability mass), signaling a limited degree of intra-distribution mobility, which implies that the European regions tended on the whole to maintain their relative positions over the study period. Overall, as in Hierro and Maza (2009) and Le Gallo (2004), the results seem to display a high level of persistence;

- the peaks of the stochastic kernel lie beneath the main diagonal, which means that the regions registered relatively greater growth rates between 1991 and 2008. This confirms the convergence process which was mentioned previously, along with the analysis of \( \beta \)-convergence;

- three peaks are well identified. Those three modes reveal a tendency towards polarization, as mentioned by Fotopoulos (2008).

Figure 4 displays the initial distribution of income per capita against its ergodic distribution for 1991. The ergodic distribution allows the identification of long-run properties of the observed distribution dynamics. Before analyzing the results, one must keep in mind that the ergodic distribution depicts what the distribution would look like in the long run if current conditions remained the same. Also, even though the ergodic distribution represents how regional income per capita develops around its mean, it does not say anything about the mean itself. When comparisons are made between the ergodic distribution and the distribution in 1991, conclusions can be drawn on the concentration/deconcentration patterns in regional income per capita, based upon current trends. Hence, assuming current trends continue, the shape of the ergodic distribution suggests that in the long run:

- it seems less probable that further improvements on convergence can be attained. The ergodic distribution is not tighter than the initial distribution, hence the following anticipation: the regions would not look more alike in the future than in the past. A closer attention to the evolution of the ergodic distribution across the two decades (see appendix, figure B) accentuates the highlighted results. This result follows Ezcurra and Rapun (2007), when they state that their evidence indicates that the convergence process detected between 1991 and 2003 will not persist indefinitely, and contrasts with the conclusions gathered by Bosker (2009) and Fischer and Stumpner (2008);

- an overall growth process is underway in Europe. The ergodic distribution presents a thicker upper tail than the actual distribution for the year at hand, hence in the future more high income per capita regions would be expected;

- there is the potential for the development of various regional clusters or convergence
clubs. The ergodic distribution does not present a `hat shape' or a `bell shape'. Instead, it displays a right skewed form with a distinctly fatter upper tail - the mass of the distribution is concentrated on the left of the figure 4 and this pattern is also visible in figure B (see appendix). Hence, one can identify an increasing presence of regions with a high income per capita, signaling more concentration in the future but over a wider array of values in comparison with the initial situation.

The evidence of a right skewed distribution for the ergodic distribution is not usual in the literature on European regions - at least, the one we are aware of. In Basile (2009), Fischer and Stumpner (2008) or Fotopoulos (2008), the ergodic distribution showed a bimodal structure; although their shape suggested that regions would look more alike in the future than in the past. Ezcurra and Rapun (2007) estimated an ergodic distribution that behaved in a quite normal way, although with a conclusion quite similar to ours, as mentioned previously. Our ergodic distribution points to a long run tendency that can accommodate more than two modes, hence we could think of a future European regional growth process with several velocities - this is an issue that deserves further attention. Are there clusters of regions in formation? Do these clusters already have different dynamic features? Moreover, do the regions that belong to the different clusters have specific characteristics? These questions should be addressed in a subsequent research.

A final set of remarks over the reliability of the results are in order. First, our analysis draws heavily on Magrini's (2007) approach, which means that the differences we are reporting could be an outcome of the differences in the estimators being used. Second, our sample is different from all the others we mentioned along the paper, both in the number of countries and in the time frame covered. Surely this fact affects the estimations and should deserve a closer attention - we are already drawing on this subject to assess the robustness of our estimates. Finally and recalling Geppert and Stephan (2008), the reduction of income disparities is a phenomenon between nation, but not between regions within the EU countries. Even if our evidence is not in line with that argument, our results might be an outcome of assuming as isolated units the various regions under study, i.e., we are ignoring the spatial characteristics of the data and the potential role of the geographical setting in shaping economic growth. Hence, this issue should deserve further attention as posit by Le Gallo (2004), Frenken and Hoekman (2006), Battisti and Di Vaio (2008) or Maza et al. (2009).

4. Conclusion

In this paper we dealt with European regional convergence in the period 1991-2008, focusing a sample of 252 regions at NUTS-2, taken from 19 European countries. In this stage of our research, our focus was not upon the validity of formal growth theories, hence we did not draw any confirmatory analysis to test a specific econometric relationship. Instead, our idea was to generate hypotheses about the underlying dynamics of the economic system, consequently the undertaken study had an exploratory nature. After covering theoretical issues related to the notions of convergence we were interested in - concepts of β-convergence and σ-convergence - we closed attention on the income distribution details. The empirical study followed the distributional dynamics approach, with the application of the methodology by Magrini (2007). Our analysis draws heavily on Magrini's proposal, namely his kernel estimator.

Recalling Capello (2008), answers for convergence problems on one hand and
endogenous determinants of regional growth on the other are needed, therefore these matters should be on the research agenda, specially on that of regional scientists. The approach proposed in this paper follows this path, since it aims to shed new light over the growth process underway in Europe. The evidence confirms some results presented elsewhere, such as the trend towards polarization or the high degree of persistence of the relative region's position and, last but not least, the admissibility of the 'convergence clubs' hypothesis.

The main novelty of our paper is the estimated shape of the ergodic distribution of the income per capita. Even though the estimated density functions are in line with those presented in the literature, and consequently corroborate the 'convergence clubs' hypothesis, the right skewed ergodic distribution of income per capita points to a long run trend towards a wider set of high income per capita regions, signaling the possible formation of new clusters, i.e. more groups of economies that present homogeneous economic growth patterns and which converge towards a common steady state. In this context, it is possible the establishment of several long run poles of attraction. As Ezcurra and Rapun (2007), our evidence also signals that the convergence process might not persist indefinitely and raises the possibility of transience in the catching-up process underway in Europe. Drawing on the exploratory study just presented, further research should be placed on the clusters of European regions on the field, namely on their dynamic features and on the characteristics of their elements.

We will end up with three final notes on the reliability of our results. First, our analysis is drawn upon a different estimator, taken from Magrini (2007). Hence our specific result on the ergodic distribution could be an outcome of our approach. However, we do not believe this, since all our results are in line with those in the literature. Second, the specific sample we use is different from all the others mentioned in the literature, both in the number of countries and in the time frame covered. We are aware of the fact that this departure inevitably impacts the estimations, and consequently we are already evaluating the robustness of the estimates to changes in the sample - including both factors, i.e. the regions and the time frame. Finally, we acknowledge the recommendations by Le Gallo (2004), Frenken and Hoekman (2006), Battisti and Di Vaio (2008) or Maza et al. (2009). Our results might be an outcome of assuming as isolated units the various regions under study, i.e., we are ignoring the spatial characteristics of the data and the potential role of the geographical setting in shaping economic growth. Hence, this issue should deserve a closer attention.

References
Bosker, M. (2009), The spatial evolution of regional GDP disparities in the 'old' and the


Andrade, C., Pinho, C., Pinho, M.F. *Exploring Regional Convergence in 19 European Countries*


Pfaffermayr, M. (2009), Conditional beta- and sigma-convergence in space: A maximum


Annex on line at the journal Website: [http://www.usc.es/economet/aeid.htm](http://www.usc.es/economet/aeid.htm)
Annex

Methodological approaches

2.1. $\beta$-convergence

The convergence hypothesis is based upon the neoclassical growth models of Solow (1956) and Swan (1956). Drawing upon the presentation by Barro and Sala-i-Martin (1999), the main features of the neoclassical growth framework can be summarized as follows:

- A closed economic system is assumed, where physical capital ($K$) and labor ($L$) are employed in the production of a homogeneous consumption good ($Y$) using a production function, which is twice differentiable, homogeneous of degree 1, increasing, jointly concave in all its arguments and strictly concave in each.

- Defining the effective amount of labor as $\bar{L} = L \cdot A_t$, where the level of technology is represented as $A_t = A_0 e^{\mu t}$ (i.e. it grows exponentially at the exogenously given rate $\mu$), and defining physical capital and output per capita of effective labor as $k = \frac{K}{\bar{L}}$ and $y = \frac{Y}{\bar{L}}$, the resulting production function including labor-augmenting technological progress is represented by $y = f(k)$.

- Under the framework of Solow and Swan, an exogenously given fraction of output is saved and invested in new physical capital while the remaining output is consumed, and the system exhibits stability, since the economy converges to a steady state equilibrium. In this position, the exogenous rate of technological progress will set the growth of the level of income per capita, the consumption per capita and the capital-labor ratio, assuming that effective labor per unit is constant.

Income convergence can be the outcome of two different processes: capital deepening; and technological catch-up. The approach of total factor productivity (TFP) convergence was pioneered by Solow and is usually known as "growth accounting". The idea beneath this approach is based on the process of technological catch-up and TFP represents the closest measure of technology. The aim of this approach is to estimate the relative portions of variation in cross-country output per worker that can be assigned to variation in factor accumulation rates and that accrues to total factor productivity. Therefore, convergence can be thwarted or accelerated depending upon the evolution over time of the gap between initial TFP differences. Evidence seems to point to differences in TFP as the main explanation of the variation in per capita income levels or in growth rates. However, this approach has specific problems, namely the usually assumed form of the aggregate production function or its inability to account for spillovers.

The term $\beta$-convergence denotes the negative correlation between the growth rate of per-capita income over a given period of time and the initial per-capita income level. If
the convergence is conceptualized in this way, the sign of the initial income variable in the so-called growth-initial level regression captures the negative correlation just mentioned. That regression is derived from the transitional dynamics of the model with Cobb-Douglas technology, through a Taylor series approximation around the deterministic steady state, which yields:

\[
\log y_t = \left( \log y^*_0 - \left( \log y^* + \log A_0 \right) \right) e^{-\beta t} + \left( \log y^* + \log A_0 + \mu \right)
\]  

where \( y^*_0 \) denotes the level of output per capita at an initial time, \( y^* \) stands for the level of output per unit of effective labor that will be reached in the steady state and \( \mu \) represents an error term that accounts for unpredictable effects on the growth rate. The average growth rate of per capita income over the period \( T \) (where \( T \geq 0 \)) starting at moment 0, is:

\[
\frac{1}{T} \log \left[ \frac{y_t}{y_0} \right] = \frac{1}{T} \left( 1 - e^{-\beta t} \right) \left( \log y^* + \log A_0 \right) + \frac{1}{T} \left( 1 - e^{-\beta t} \right) \log A_0 + \mu
\]  

Hence, ceteris paribus, the average growth rate of per capita income is negatively related with its initial level, and it is conditioned on the exogenous growth rate of technology, on the steady state value of income per effective worker, and on the initial level of technology. Since the conditioning factors have an unknown form, the so-called growth-initial level regression usually has the following specification proposed by Barro and Sala-i-Martin (1991, 1992):

\[
\frac{1}{T} \log \left[ \frac{y_t}{y_0} \right] = \alpha_0 + \lambda_0 \log y_t + \mu
\]  

where \( y_{i,t} \) stands for per capita income of economy \( i \) (\( i=1,..., N \)) at the moment \( t \), \( t \) is the time length and \( \mu_t \) represents an error term that accounts for unpredictable effects on the growth rate.

It must be noted that:

\[
\alpha_0 = g + \frac{1}{T} \left( 1 - e^{-\beta t} \right) \left( \log y^* - \log A_0 \right)
\]  

\[
\lambda_0 = \frac{1}{T} \left( 1 - e^{-\beta t} \right)
\]  

\[
\beta = (1 - \gamma) (n + g + \delta)
\]  

\[
\text{half-life} = \frac{\log 2}{\beta}
\]  

where \( g \) stands for the rate of technological change, \( y^* \) denotes the steady state income level, \( A_0 \) is the initial technological level, \( \gamma \) the income elasticity to capital, \( n \) the population growth rate and \( \delta \) the capital depreciation rate. \( \beta \) has been known in the literature as the rate of convergence, since it gives the mean speed at which the gap
Andrade, C., Pinho, C., Pinho, M.F.  *Exploring Regional Convergence in 19 European Countries*

between the steady state and its current level is being closed. "Half-life" is a measure of the time the representative economy needs to halve the gap between initial and steady state levels.

The introduction of the $\beta$-convergence concept drives to an important conceptual distinction: there is a need to distinguish unconditional (or absolute) convergence from conditional convergence. From the above analysis, it is possible to derive a function for the steady state income level that shows its dependence on $g$, $δ$, $n$, $γ$, $A_0$, and also upon the investment rate (for easiness, $\theta$ is a vector that combines these parameters). If all elements in $\theta$ are the same for all economies considered, the target notion of convergence will be that of unconditional (or absolute) convergence. Hence, the economies under study would have similar structures, i.e., identical features for consumer preferences, technologies, economic policies, as well as technological change and savings.

The concept of conditional convergence generalizes the absolute convergence approach by allowing differences in the steady states across economies. The idea is that independent economies will display diminishing initial income differences only if the steady states differ less than the initial incomes. The possibility of income gaps arises when that does not happen - this is likely since each economy converges independently from the others to its own steady state. This approach requires an appropriate set of variables to be included in the previous equation 3 in order to control the differences - see equation 8.

$$\frac{1}{T} \log \left[ \frac{y_i}{y_0} \right] = \alpha_0 + \lambda_0 \log y_i + \psi X + \mu$$

(8)

where $X$ stands for the set of explanatory variables that are the steady state determinants. For instance, Karras (2008) presents empirical findings that clearly reject absolute convergence in income per capita but are very strongly supportive of conditional convergence.

One must notice that $X$ represents those growth determinants that lie outside Solow and Swan's approach (Durlauf et al. 2005). Therefore, standard neoclassical growth models imply unconditional convergence. Withal, equation 8 represents the baseline for much of growth econometrics and, due to the lack of theoretical support; the literature unfolded a large number of potential growth regressors and an intense debate over their appropriateness. Public consumption, public investment or human capital are variables commonly used in the literature. But other proxies are being reported as well (Durlauf et al. 2008), such as economic institutions, legal and political systems, climate, geographic isolation, ethnic fractionalization, or culture.

The $\beta$-convergence concept encompasses another relevant dimension for the understanding of convergence. We already covered the notions of unconditional convergence and conditional convergence. The first one implies one equilibrium level to which all economies close in. The latter involves different equilibria for each economy, since each one approaches its own and unique equilibrium. Hence, there might be space in this second view to conceive the possibility of a group of countries approaching a
particular equilibrium if they share initial attributes that correspond to that equilibrium. In this case, the sense for the term convergence is that of club (or local) convergence and the model should yield multiple equilibria.

2.2. $\sigma$ -convergence

An alternative strategy for the empirical research is to focus upon the distribution of incomes as the unit of analysis, instead of individual economies themselves. Hence, the attention would be on the shape of the income distribution as a whole, rather than on the relative locations within the distribution.

Several methods have been employed to the distribution's analysis. Initially, the research focused on how the dispersion of income differences developed over time. This notion of economic change, known as $\sigma$-convergence, occurs if the cross-sectional standard deviation of per capita income ($\sigma$) is falling over time. Hence:

$$\sigma_{t+1} - \sigma_t < 0 \quad (9)$$

This notion continues to deserve attention until now, as it is shown by Villaverde (2006) and Young et al. (2008). Villaverde (2006) focuses on the Spanish regions to conclude that the high level of persistence justifies the need for a regional policy. The latter is centered on the so-called Gallon's fallacy, which states that $\beta$-convergence implies $\sigma$-convergence. The work of Young et al. (2008) adds new evidence to this discussion, based upon a very large dataset. The fallacy has been proved erroneous, however, as these authors note once again; $\beta$-convergence is a necessary but not sufficient condition for $\sigma$-convergence.

Unlike $\beta$-convergence, statistical tests for $\sigma$-convergence were not readily available. An avenue of research has been focused on identifying regression specifications to build inferences for $\sigma$-convergence. However, great difficulties can be pointed out in interpreting tests of $\sigma$-convergence, since they assume that the data generating process is not invariant. Since evidence points to evolving distributions for the data, there is uneasiness about test distributions under the null hypothesis of $\sigma$-convergence.

Another issue that has deserved some recent attention is the formulation of a sensible notion of conditional $\sigma$-convergence. Durlauf et al. (2005) calls attention for the need of such a contribution. Pfaffermayr (2009) defines conditional $\sigma$-convergence as the average variance reduction of the predicted per capita income and introduces Wald tests for the conditional $\sigma$-convergence while using a maximum likelihood approach.

2.3. Income distribution dynamics

Although intuitively simple, $\sigma$-convergence has some limitations: it ignores other aspects of the income distribution, such as skewness and modality; and it masks churning processes and the mobility of individual economies, due to the undertaken aggregated approach. Therefore, more sophisticated approaches arose in the literature. An active area has been the estimation of income densities with a variety of parametric and nonparametric methods. The approach was introduced by Quah during the 90s and uncovered two important stylized facts that need explanation from growth theory: there is persistence, i.e., most economies remain in the same position of the distribution; there is bimodality, i.e. as time goes by the middle range of the distribution gets thinner and the two tails pile up (see Quah 1996, 1997 for further details).
One way to study the features of income distributions is by estimating income densities with stochastic kernels and by analyzing their evolution over time. Rey and Janikas (2005) refer the use of stochastic kernels to study changes in modality, distributional mixing and stochastic dominance. Recently, Maza et al. (2009) also apply stochastic kernels, although with a variant that was suggested by Hyndman, Bashtannyk and Grunwald, in 1996. This new approach is based on the estimation of the stacked density plot and highest density plot and aims to find variables that can partially explain differences in productivity whose nature was previously found persistent. One of the main results of the paper is that space matters, as well as investment per employee and compensation per employee.

Another path for the study of the income distributions features is through Markov chain modeling. This technique can be viewed as a discretization of the income distribution into nonoverlapping income intervals, that will be used to estimate the probability for the transition of an economy that was on as interval to another one (or the probability for staying on the same interval) in a future time period. As pointed out by Rey and Janikas (2005), this methodology allows a new approach to the dynamics of the income distribution features, such as the ergodic steady state income distribution, the transitional speed to the steady state distribution, the convergence and/or polarization tendencies and the identification of measures of intra-distributional rank and class mobility. Quah presented an alternative formulation designed to avoid discretization problems, since this methodology can arbitrarily alter the probabilistic properties of the data simply by changing the discretization scheme. Supposing a group of \( n \) economies and assuming that time \( t \) is continuous, the level of per capita income of economy \( i \) at time \( t, \ y_i(t) \), can be transformed through its normalization with respect to the average level of per capita income for the set of economies under study at the same point in time (which is represented by \( \bar{y}(t) \)). Hence, the transformation is:

\[
x_i(t) = \frac{y_i(t)}{\bar{y}(t)}, \text{ with } x_i(t) \in l
\]  

(10)

Assuming a stochastic process \( \{x(t) : t \geq 0\} \) and denoting by \( F_{x(t)} \) the distribution of \( x(t) \) and by \( f_{x(t)} \) a probability measure associated to \( F_{x(t)} \), the dynamics of the cross-sectional distributions of per capita income can be modeled as an autoregression such as:

\[
\forall A \subset l : f_{x(t+s)}(A) = \int_{-\infty}^{\infty} M_{t,s}(x,A)f_{x(t)}(x)dx
\]

(11)

where \( M_{t,s} \) is a stochastic kernel that maps the density at time \( t \) into the density at time \( t+s \), tracking where points in \( f_{x(t)} \) end up in \( f_{x(t+s)} \). It is also assumed that the sequence \( \{M_{t,s} : t \geq 0\} \) is time invariant and that stochastic process is Markov. The operator \( M_{t,s} \) can be estimated through nonparametric methods.

Carrington (2006) used Markov chain as a starting point and departed from the literature on convergence by applying second-order stochastic dominance (SOSD) analyses. Carrington claims that SOSD overcomes the disadvantages of Markov chain
modeling, namely the distortions brought by the discretization. Moreover, according to Carrington, SOSD allows more links between economy theory and statistical measurement, although these relations are better suited for utility and choice problems than growth issues. For each pair of sets of income per capita $X_1$ and $X_2$, with cumulative distributions functions $F_1$ and $F_2$, the null hypothesis to test\(^3\) is:

$$H_0 : \int_{-\infty}^{+\infty} F_1(u) du \leq \int_{-\infty}^{+\infty} F_2(u) du \text{ for some } x \in [a, b]$$  \hspace{1cm} (12)

against the alternative:

$$H_1 : \int_{-\infty}^{+\infty} F_1(u) du > \int_{-\infty}^{+\infty} F_2(u) du \text{ for some } x \in [a, b]$$  \hspace{1cm} (13)

where $[a, b]$ is the range of $X_1$ and $X_2$. One must note that this strategy also has its shortcomings, especially since it does not offer information over the magnitude and causes of the identified changes as well as it lacks an indication of the speed of change that the test is designed to detect.

Rey and Janikas (2005) mention a final approach to distributional modeling, which is based on regression trees. Labeled as a less ad hoc methodology, this approach allows the identification of convergence clubs as well as parameter heterogeneity in the underlying growth models.

A common critic pointed out to all the strategies for exploratory analyses of convergence is their disregarding for the implications of space. Spatial dependence has just begun to attract attention, as it is shown by Maza et al. (2009).

**Appendix: figures**

**Figure A: Estimated density functions**

---

\(^3\) The details of the test statistic are presented in the paper by Carrington (2006).
Figure A3: 2000 and 2005 ln GDPpc

Figure A4: 2005 and 2008 ln GDPpc

Figure B: Estimated ergodic density and initial distribution of regional income per capita

Figure B1: Estimated ergodic density and initial distribution of regional income per capita in 1991-1995

Figure B2: Estimated ergodic density and initial distribution of regional income per capita in 1995-2000

Figure B3: Estimated ergodic density and initial distribution of regional income per capita in 2000-2005

Figure B4: Estimated ergodic density and initial distribution of regional income per capita in 2005-2008