M. SIMONA ANDREANO, ROBERTO BENEDETTI,
PAOLO POSTIGLIONE

HETEROGENEITY IN THE ECONOMIC GROWTH OF EUROPEAN REGIONS: AN ADAPTIVE GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH

Universitas
Mercatorum

Università telematica delle
Camere di Commercio Italiane

Novembre 2010

HETEROGENEITY IN THE ECONOMIC GROWTH OF EUROPEAN REGIONS: AN ADAPTIVE GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH

M. SIMONA ANDREANO[a], ROBERTO BENEDETTI[b]
AND PAOLO POSTIGLIONE[b]

[a] *Faculty of Economics, Universitas Mercatorum, Rome, Italy*
[b]*Department of Business, Statistical, Technological and Environmental Sciences,*
University of Chieti-Pescara, Italy

I[a] edizione novembre 2010

Realizzazione editoriale: Università Telematica Universitas Mercatorum, Roma.
Disponibile online nel novembre 2010
Curata da Università Telematica Universitas Mercatorum, Roma.

ISBN 978-88-905081-4-1
ISBN 978-88-905081-4-1-A

SUMMARY

This paper proposes a new technique based on an adaptive weighted regression procedure to verify the presence of convergence clubs in a cross-section of regions. The approach extends a procedure originally proposed in the field of image analysis based on the assumption of local homogeneity of the signal. The presence of the heterogeneity is a criterion to divide the sample of observations (regions) into smaller homogenous groups (clubs). Our results highlight the presence of five different clubs with different laws of motion for growth within each subgroup. Spatial dependence is also considered in the definition of the economic convergence model.

KEYWORDS: convergence clubs, local stationarity, weighted regression, spatial dependence.

1. INTRODUCTION

In the past few years the economic convergence hypothesis has been the subject of intense debate (Baumol 1986, Barro 1991, Mankiw et al. 1992). The controversy has been largely empirical, focusing on three different competing hypotheses: absolute, conditional and convergence clubs.

Empirical evidence on convergence of national economies has usually been investigated by regressing growth rates of real Gross Domestic Product (GDP) on initial levels, sometimes after correcting for some exogenous variables. A negative regression coefficient (the β-coefficient) is interpreted as an evidence of convergence, as it implies that, on average, countries with low per-capita initial GDP are growing faster than those with high initial per-capita GDP. Moreover the conditional convergence hypothesis suggests that among countries that are similar in preferences, technologies, population growth rates, political stability, the lower the levels of output per capita the higher the growth rates. Thus, countries that are similar except for their initial conditions are expected to converge to the same steady-state and hence to one another (Barro and Sala-i-Martin 1995).

The question why some countries grow faster than others has been the focus of the extensive empirical literature on conditional convergence. Mankiw et al. (1992), Barro and Sala-i-Martin (1992), Islam (1995), Canova and Marcet (1995), and Caselli et al. (1996), among the others, find evidence that countries are converging to their individual balanced growth paths. Despite the availability of many complementary and alternative definitions of convergence, economists seem to agree that unconditional convergence is not realized in a sample of cross-country regressions. This led Baumol (1986) to suggest that economic convergence could be achieved if we consider groups of countries, within which we observe convergence, but that do not converge to each other. This hypothesis is known as convergence clubs (Galor 1996). Countries that are identical in their structural characteristics but differ in their initial level or distribution of human capital may cluster around different steady-state equilibria in the presence of i.e. social increasing returns to scale from human capital accumulation (e.g. Azariadis and Drazen 1990), capital market imperfections (Galor and Zeira 1993), parental and local effects in human capital formation (Durlauf 1996), imperfect information (Tsiddon 1992) and non-convex production function of human capital (Becker et al. 1990).

The issue of convergence clubs requires the application of non-standard econometric techniques that allow us to divide the whole sample into smaller groups. Instead of assuming that the regression coefficients are the same for all units belonging to one group, it allows for a further layer of heterogeneity within groups.

Two main approaches are used to verify the presence of convergence clubs. The first selects the composition of the potential convergence clubs according to some externalities, such as threshold or discriminant variables and often applies cluster analysis techniques. The second focused on the knowledge of the distribution and density of the per-capita growth rate.

In the first class we find Durlauf and Johnson (1995), who study cross-country heterogeneity, providing evidence that there are multiple poles of attraction in the growth process. They use a regression-tree procedure to determine threshold levels of initial per-capita GDP and literacy rates, which imply groups of countries satisfying common linear cross-sectional regression equations. Azariadis and Drazen (1990), Feve and LePen (2000), and Ramajo et al. (2008) apply threshold switching procedures to identify convergence clubs relating to the determination of the clusters thresholds to critical level in i.e. the accumulation of physical and human capital. Desdoigts (1999) investigates the sources of heterogeneity across a worldwide set of countries by using the Exploratory Projection Pursuit (EPP) data analysis. Postiglione et al. (2010) propose a modified regression tree algorithm for both the classical and the spatial β-convergence model in order to identify convergence clubs in European regions. The recent Unified Growth Theory (Galor 2007) also suggests that the presence of various convergence clubs is linked primarily with critical changes in the rates of technological progress, population growth and human capital formation.

Referring to the second approach for the determination of convergence clubs, Quah (1996, 1997), Bianchi (1997) and Canova (2004) estimate the distribution density function of the data allowing for heterogeneity within groups, where equilibria may display multiple basins of attraction. Quah (1997) has analyzed patterns of economic growth across countries from the perspective of distribution dynamics, considering the distribution of GDP across countries. Thus, the problem considered in the Quah's study differs from the classical set of questions prominently considered by many authors (e.g. Galor and Zeira 1993). He found that the distributions weighted by the relative numbers of people in each economy present a profound empirical regularity: an "emerging twin peaks" cross-sectional distribution. Bianchi (1997) considers bootstrap non-parametric techniques for the estimation of the density distribution of GDP across countries and proposes a multimodality tests for intra-distribution dynamics. Canova (1999, 2004) applies an Empirical Bayesian method based on the distribution density of per-capita GDP and concludes that European regional data clusters around four poles of attraction with different economic features.

The main purpose of this paper is to introduce a new method for the identification of convergence clubs. Our approach refers to a technique first described by Polzehl and Spokoiny (2000) for the image segmentation and extended here for the estimation of local homogeneous geographical zones that can be interpreted as convergence clubs. The approach is general and uses an adaptive weighted regression procedure.

The starting point of the analysis is the hypothesis that there may be significant heterogeneities in the cross-sectional data and natural clustering of geographical units around different poles of attraction. To study cross-country growth behaviour in favour of multiple regime alternatives, we use geographically weighted regression. The use of geographically weighted regression provides estimates of coefficients for each variable and each geographical location. Similarity in the local estimated models will suggest the aggregation of the corresponding regions into the same group. The procedure is applied iteratively. Thus heterogeneity is a criterion to divide the whole sample of regions into smaller homogeneous clubs.

Many features distinguish the proposed approach from existing ones. First the identification of groups is made without considering some control variables or external information, but is data-driven. The absence of hypothesis on the number of clubs represents an important feature of our technique.

The use of geographically weighted regression allows to implicitly incorporating non-stationarity in the space in the modelling process. Furthermore we focus on the presence of spatial dependence and on its role in the identification of convergence clubs, as evidenced by many authors (e.g. Ramajo et al. 2007, Ertur and Koch 2007).

We use the neoclassical Solow growth model as a general framework to analyze economic convergence across the 187 NUTS 2 EU regions, for the period 1981-2004. In light of the next Eastern enlargement of the European Union (EU), it is interesting to look deeper into the economic progress of past accession candidates and to verify if the EU integration conduces to convergence through EU regions, to evaluate the effectiveness of the Cohesion Policies.

The paper is organized as follows. Section 2 reviews the economic convergence models describing both the classical and the spatial one. In section 3 we briefly present the geographical weighted regression and our proposed adaptive procedure to identify convergence clubs. Section 4 uses the adaptive weighted regression procedure to identify the clubs of European regions and summaries the empirical results. Section 5 concludes.

2. THE ECONOMETRIC MODEL

The starting point of our analysis is the augmented Solow-Swan growth model assumed by Mankiw et al. (1992), hereafter MRW. In their paper the authors included accumulation of human capital as well as physical capital, to provide a more complete explanation of why some countries are rich and other poor. Hence, the Cobb-Douglas production function at time t becomes:

$$Y_t = C_t^u H_t^v (A_t L_t)^{1-u-v} \tag{1}$$

where Y is the output, C is the physical capital stock, H is the human capital stock, L is the labour, A is the technology level and, u and v are constants, with $u + v < 1$. The technology term is assumed to increase exponentially, $A_t = A_0 e^{xt}$ with constant rate x and each country augments its physical and human capital stocks at the constant saving rate s_c and s_h, while both stocks depreciate at the same rate δ.

The Cobb-Douglas function can be written in its *intensive form*, with quantities effective unit per-worker. The conditions for the changes in the capital-labour ratio c and human-labour ratio h over time are:

$$\begin{aligned} \dot{c} &= s_c \cdot y - (x + n + \delta)c \\ \dot{h} &= s_h \cdot y - (x + n + \delta)h \end{aligned} \tag{2}$$

where a dot over a variable denotes the derivative with respect to time, $\delta > 0$ and $n = \dot{L}/L$.

Solving equation (2) we obtain:

$$c^* = \left[\frac{s_c^{1-v} s_h^v}{x+n+\delta} \right]^{1/(1-u-v)}$$
$$h^* = \left[\frac{s_c^u s_h^{1-u}}{x+n+\delta} \right]^{1/(1-u-v)} \qquad (3)$$

Therefore the steady-state of capital-labour ratio is related directly to the rate of saving and inversely to the rate of population growth. Substituting equation (3) into the production function expressed in its intensive form, taking logs and subtracting per-worker initial GDP, the growth of per-worker output between period 0 and *t* is:

$$\ln(y_t) - \ln(y_0) = \theta \frac{u}{1-u-v} \ln(s_c) + \theta \frac{u}{1-u-v} \ln(s_h) - \theta \frac{u+v}{1-u-v} \ln(x+n+\delta) - \theta \ln(y_0) \qquad (4)$$

where $\theta = 1 - e^{-\lambda t}$ and λ is the rate of convergence. From equation (4) we can see that the presence of human-capital accumulation increases the impact of physical capital accumulation on per-worker GDP.

The model estimated in our application, for each region *i*, is expressed in an unrestricted form as in MRW framework:

$$\ln\left(\frac{y_{it}}{y_{i0}} \right) = \alpha + \beta \ln(y_{i0}) + \pi_1 \ln(s_{ic}) + \pi_2 \ln(x+n+\delta)_i + \pi_3 \ln(s_{ih}) + \varepsilon_i \qquad (5)$$

where $\varepsilon_i \propto N(0, \sigma_\varepsilon^2)$ is the disturbance term.

The conditional β-convergence hypothesis is verified if $\beta<0$ is statistically significant. The MRW model says that differences in saving, human capital and population growth should explain cross-country differences in per-worker GDP.

However, in our analysis, we want to see if this model is consistent with the presence of multiple steady-state equilibria that classify regions into different groups with different convergence characteristics. The identification of such multiple regimes and the evidence of convergence clubs take explicitly into account for parameter heterogeneity. To detect distinct poles of attraction we estimate the model with adaptive geographically weighted regression method (henceforth AGWR). A brief review of the basic concepts of classical GWR and its modified adaptive version will be presented in the next section.

Finally, we are aware that spatial externalities involve technological interdependence among countries, therefore models of economic growth need to include spatial proximity (contiguity) effects, associated, for example, with localized knowledge spillovers and interfirm demand-supply networks. Various tests to detect the presence of spatial effects in the estimation of the appropriate β-convergence model have been described in Anselin et al. (1996). To provide empirical evidence of spatial effects on growth, the MRW model for the identification of convergence clubs can be extended to incorporate spatial dependence. The *Spatial Autocorrelation Model* (*SAM*) introduces an autoregressive term in the model as:

$$\ln\left(\frac{y_{it}}{y_{i0}}\right) = \alpha + \rho D \ln\left(\frac{y_{it}}{y_{i0}}\right) + \beta \ln(y_{i0}) + \pi_1 \ln(s_{ic}) + \pi_2 \ln(x+n+\delta)_i + \pi_3 \ln(s_{ih}) + \varepsilon_i \quad (6)$$

where *D,* the spatial weight matrix, and ρ is the spatial autoregressive coefficient.

An alternative way to include spatial effects is represented by the *Spatial Error Model* (*SEM*), where spatial dependence is taken into account through the error term, that follows an autoregressive scheme:

$$\varepsilon_i = \varphi D \varepsilon_i + u_i \qquad u_i \propto \quad N(0, \sigma_u) \tag{7}$$

In the SEM model φ determine the degree to which the values of individual locations depend on their neighbours.

3. THE ALGORITHM FOR THE IDENTIFICATION OF CONVERGENCE CLUBS

The identification of convergence clubs is obtained through a partition of an area into groups of geographical zones not necessarily conterminous that have similar growth path. In this article, in order to identify these geographical clusters we use a modified adaptive version of GWR.

The GWR extends the traditional regression framework by allowing local rather than global parameters to be estimated. The general regression model can be rewritten as:

$$y_i = b_0(i) + \sum_h b_h(i) x_{ih} + \varepsilon_i \tag{8}$$

where i=1,…,n is the geographical region on the observed space, and x_{ih}, h=1,.., H are the explanatory variables. Thus GWR equation recognizes that spatial heterogeneity in the model might exist and provides a way to measured it.

As evidenced by Fotheringam et al. (2002), observed data that are contiguous to region i have more influence in the estimation of the $b_h(i)$s than data located farther from i. In GWR an observation is weighted in accordance with its proximity to location i and the vector estimate of $b_h(i)$, $\hat{\mathbf{b}}(i)$, is given by:

$$\hat{\mathbf{b}}(i) = (\mathbf{X}'\mathbf{W}(i)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(i)\mathbf{y} \tag{9}$$

where $\hat{\mathbf{b}}(i)$ is the locally stationary parameter vector and $\mathbf{W}(i)$ is an $n \times n$ matrix whose off-diagonal elements are zero and diagonal elements denote the weights of each of the n observed data for regression region i and is defined as:

$$\mathbf{W}(i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & w_{in} \end{bmatrix} \tag{10}$$

The equation (9) represents a weighted least squares estimator but rather than having a constant weight matrix, the weights in GWR vary according to the region i. To overcome the problem of overparametrization, we need to express the weights w_{ij} as a Kernel function of d_{ij}, the distance between region i and region j: $w_{ij} = K_d(d_{ij})$.

Generally, desirable features of a Kernel function K_d are:

a) $K_d(0) = 1$

b) $\lim_{d \to \infty} K_d(d) = 0$ (11)

c) K_d is a monotone decreasing function for positive real numbers.

Expression (9) can be written for $i = 1,\dots,n$ as a matrix, with each $\hat{\mathbf{b}}(i)$ corresponding to a column whose elements are $\hat{b}_h(i)$. In this way it is possible to see how a coefficient corresponding to a given explanatory variable changes geographically.

In this paper we extend the GWR iteratively, by computing at each iteration new weights in (10) by comparing the estimated coefficients $\hat{\mathbf{b}}(i)$ over all the regions $i = 1,\dots, n$ and aggregating in a group regions that present similar estimates.

This adaptive weighted regression procedure (AWR) was originally introduced by Polzehl and Spokoiny (2000) in the context of image denoising and was generalized to the case of an arbitrary local linear parametric structure (Polzehl and Spokoiny 2003a, 2003b).

The idea was to estimate a constant function $f(\mathbf{X}_i;\mathbf{b}(i)) = \mathbf{b}(i)$ separately at each design point i by using locally constant modelling with an adaptive choice of the neighbourhood in which the applied model fits the data well. The n design points were partitioned into M disjoints heterogeneous groups $A_1,..,A_M$ and the function is constant within each group A_m. The adaptive procedure suggested how to recover for every point i the corresponding group A_m by analyzing the similarity of the estimated local constant models. The number of groups, the difference between values of the image function f

for different groups and the regularity of edges are unknown and may be different for different parts of the study area.

The problem is similar to the aim of the present paper, where the function is the MRW model and the image design points are the *n* regions. Therefore we combine the GWR with the AWR, modifying some rule in the procedure and apply the new adaptive GWR technique (hereafter AGWR) on the EU-NUTS2 regions to investigate the presence of convergence clubs.

For the initial step, the estimates $\hat{\mathbf{b}}^{(0)}(i)$ are computed for each region $i = 1,\dots,n$ applying (9), where $\mathbf{X}(i)$ is the vector of explanatory variables in the MRW model (5). The starting weights are:

$$w_{ij}^{0} = K_d(d_{ij}) = e^{-\gamma d_{ij}} \tag{12}$$

where the exponential function was chosen for the Kernel function according to Fotheringam et al. (2002).

Polzehl and Spokoiny (2003b) define the location Kernel K_d for the choice of the weights fixing a bandwidth, and computing in each iteration the distances d_{ij} for all points included in the neighbourhood $U(i)$ of the point *i*. We do not impose such a restriction and take into account all the regions $j \neq i$.

In the next steps the weights w_{ij} in $\mathbf{W}(i)$ are iteratively calculated to assign each region *i* to a locally homogeneous zone. We aim to develop a data driven rule to evaluate whether the local model corresponding to region *j* is not different from the model at region *i* and therefore to define new weights w_{ij} that assign region *i* and *j* to the same local model in the next algorithm's iteration. At each iteration *l* we compute updated weights $w^l(i,j)$ defined by comparing the estimates $\hat{\mathbf{b}}^{l-1}(i)$ and $\hat{\mathbf{b}}^{l-1}(j)$ obtained in the previous iteration; if the statistic $\hat{\mathbf{b}}^{l-1}(i) - \hat{\mathbf{b}}^{l-1}(j)$ is large compared to its standard deviation, then these two regions *i* and *j* are classified in different groups. These weights are then used to compute new improved estimates. Therefore we propose to compute the statistics T^l_{ij}:

$$T^l_{ij} = \left(\hat{\mathbf{b}}^{l-1}(i) - \hat{\mathbf{b}}^{l-1}(j)\right)\mathbf{\Sigma}^{-1}\left(\hat{\mathbf{b}}^{l-1}(i) - \hat{\mathbf{b}}^{l-1}(j)\right)^t \tag{13}$$

where $\mathbf{\Sigma}$ is the pooled estimator of the variance (i.e. Casella and Berger, 1996) and *t* indicates the transpose. The value of this statistics is considered as a penalty, that is, the new weights at the next iteration step w^l_{ij} is penalized by large values of T^l_{ij}.

In some cases including even one region *j* into the new local model of region *i* may significantly change the estimate $\hat{\mathbf{b}}^{l-1}(i)$. To prevent from this change we need to adjust K_d for the penality for extending the model in each iteration. This control step guarantees that further iterations do not lead to an essential decrease in the accuracy estimation. In that situation, neither the statistical penalty nor the penalty for extending the model would significantly affect the estimates obtained after the first $l - 1$ iterations. Therefore at each iteration step the location penalty K_d is relaxed by increasing *l* at cost of introducing a data-driven statistical penalty. Polzehl and Spokoiny (2000) increased at each iteration *l* the radius h_l in the neighbourhood $U(i)$ of region *i* and they recalculated the estimates over a larger neighbourhood.

Finally we can define the new weights w^l_{ij} as the product of the Kernel functions on $d^l_{ij} = {}^{d_{ij}}\!/_{l}$ and on T^l_{ij}:

$$w_{ij}^{l} = K_d(d_{ij}^{l}) \cdot K_T(T_{ij}^{l}) = \left(e^{-\gamma d_{ij}/l}\right) \cdot \left(e^{-\tau T_{ij}^{l}}\right) \tag{14}$$

Another useful parameter to control the algorithm behavior is η, that stabilizes the AGWR procedure with respect to the iterations, comparing the new weight w_{ij}^{l} in (14) on the value obtained at the previous step:

$$\dot{w}_{ij}^{l} = (1-\eta)\dot{w}_{ij}^{l-1} + \eta w_{ij}^{l} \tag{15}$$

We can now summarize the iterative procedure as follows:

1. Initialization: for each region i calculate initial estimates $\hat{\mathbf{b}}^{0}(i)$ with standard GWR, where $w_{ij}^{0} = K_d(d_{ij}) = e^{-\gamma d_{ij}}$ and compute $\hat{\sigma}_{\varepsilon}^{2}$;
2. Computation of the weights $\dot{w}_{ij}^{l}$: at each iteration l compute the statistics $T_{ij}^{l} = \left(\hat{\mathbf{b}}^{l-1}(i) - \hat{\mathbf{b}}^{l-1}(j)\right)\Sigma^{-1}\left(\hat{\mathbf{b}}^{l-1}(i) - \hat{\mathbf{b}}^{l-1}(j)\right)^{t}$, $K_d(d^{l}_{ij})$, $K_T(T^{l}_{ij})$ and determine $w_{ij}^{l} = K_d(d_{ij}^{l}) \cdot K_T(T_{ij}^{l}) = \left(e^{-\gamma d_{ij}/l}\right) \cdot \left(e^{-\tau T_{ij}^{l}}\right)$. Apply the convex combination to determine the new weights $\dot{w}_{ij}^{l}$ that represent the diagonal elements of $\mathbf{W}^{l}(i)$.
3. Stopping: If $\max\left|\dot{w}_{ij}^{l-1} - \dot{w}_{ij}^{l}\right| < \omega, \ \forall i,j$ with ω a fixed small value, the procedure is stopped. Then use the current weights to estimate the final model with GWR. Otherwise continue.
4. Estimation of the new model: Use the new matrix $\mathbf{W}^{l}(i)$ to estimate the model with GWR and return to step 2.

The main differences of our procedure with respect to that proposed by Polzehl and al. (2000) rely in the determination of $K_d(d^l_{ij})$ and $K(T^l_{ij})$, in addition to the regression model considered in $f(\mathbf{X}_i;\mathbf{b}(i))$.

The choice of the parameters γ, τ and η that enter in the computation of the weights are crucial for the execution of the AGWR procedure. The memory parameter η is taken by default 0.5. The choice of the other two parameters is data dependent and will be presented in the next section.

If the AGWR procedure is applied on the MRW model extended in its spatial SAR or SEM formulation, we refer in the local homogeneous model to equations (6) and (7). The estimation of the vector $\mathbf{b}(i)$ will change consequently, using maximum likelihood or GMM, as proposed by Cressie (1993).

4. THE EMPIRICAL RESULTS

The data used in our empirical analysis to identify convergence clubs in European Union consist in 187 NUTS 2 regions, spanning the period 1981-2004, of 12 countries (Austria, Belgium, Finland, France, Germany, Greece, Italy, Portugal, Spain, Sweden, the Netherlands, and the United Kingdom). Annual data were obtained from two different sources: Eurostat REGIO for the human capital and Cambridge Econometrics data set for all the other variables.

The dependent variable is the natural logarithm of per-worker GDP growth rate and the conditioning variables are: the saving rate (δ), the population growth (n), the level of technology growth rate (x), the depreciation rate of capital (s_c) and the human capital (s_h), referring to MRW framework. The saving rate δ is measured as the average of the investments in percentage of GDP over the period 1981 to 2004. The human capital (s_h) is the adult literacy rates defined as the fraction of population on age 25-64 that has the highest education level (ISCED level 3-4). The spatial weight matrix D is defined in terms of normalized distance from the five nearest neighbour's regions.[1]

Estimates for the MRW model over the whole sample in its standard and spatial (SAR and SEM) versions are presented in Table I. The estimates of the parameter relating β-convergence are in all three models negative and highly statistically significant, confirming the presence of conditional convergence. The behaviour of the models is quite similar, showing a slow speed of convergence (λ), a not significant human capital variable and an unsatisfactory value of R^2. The time necessary for the economies to fill half of the variation with respect to the steady state is over 98 years for the OLS and the SEM models.

Table I. Estimates of models for global convergence

	α	β	π_1	π_2	π_3	λ	R^2	*Spatial Parameter*
OLS	0.040 (0.000)	-0.007 (0.000)	-0.013 (0.028)	-0.002 (0.003)	-0.003 (0.568)	0.647%	0.183	-
SAR	0.031 (0.000)	-0.006 (0.000)	-0.010 (0.075)	-0.001 (0.018)	-0.001 (0.845)	0.561%	0.265	0.329 (0.000)
SEM	0.041 (0.000)	-0.007 (0.000)	-0.013 (0.038)	-0.001 (0.025)	-0.004 (0.403)	0.647%	0.273	0.365 (0.000)

p-values are reported in parenthesis

[1] Euclidean and Contiguity distance matrices were also considered, but methodological considerations and empirical results suggest to choose the normalized Contiguity one.

The presence of spatial effects is highlighted by significant spatial parameters ρ and φ in the SAR and SEM models. In order to identify the form of the spatial dependence, the LM tests and their robust version are performed on the OLS model and the results appear in Table II. Following the decision rule suggested by Anselin and Florax (1995), the most appropriate specification is the spatial error model.

Economic convergence in the whole sample could not hold or be weak because countries belonging to different regimes. Therefore we attempt to identify the presence of multiple regimes in the data through the use of the AGWR algorithm presented in the previous section and to check whether convergence holds within these clubs. In this way we also expect to improve our estimated models.

Table II. Diagnostics for Spatial Dependence

Spatial Test	Statistics
Moran's I	0.231 (0.000)
LM (Lag)	21.087 (0.000)
LM (Error)	25.006 (0.000)
Robust LM(Lag)	0.009 (0.926)
Robust LM(Error)	3.928 (0.047)

p-values are reported in parenthesis

Before starting with the AGWR we need to choose the parameters γ and τ that enter in the computation of the weights $\dot{w}_{ij}^{l}$ in (15). Large values of γ and τ discourage the aggregation of the regions in clubs and get low sensitivity to local heterogeneity, while low levels lead to over penalization and to unstable performance. According to Herault

(2000) and by computing different data-driven simulation we adopt $0.1 \leq \gamma \leq 0.3$ and $0.1 \leq \tau \leq 0.5$.

The presence of spatial autocorrelation in the error terms can invalidate the inferential basis of the test methodology of OLS model, however OLS estimates remain unbiased. Given the unbiasedness and the low computational complexity of the OLS model, we apply the AGWR procedure also on it. Therefore we perform the AGWR with both, the OLS and the SEM model. We also applied the AGWR procedure on the SAR model, but we do not report the results to save space; moreover in this case the club identification and the estimated models were not satisfactory.

Table III contains some summary statistics applied on the clubs obtained by the AGWR procedure, for $\gamma \in [0.1; 0.3]$ and $\tau \in [0.1; 0.5]$, with a grid search of 0.1.

The statistics reported in Table III are the R^2, the number of clubs, the number of points left out from the estimation, the minimum and the maximum number of regions present in the clubs determined by the procedure, for different combinations of γ and τ and for both estimated models. It is important to highlight that if a club does not reach a minimum number of observations, the model cannot be estimated; therefore these regions will be excluded from the analysis. The statistic *number out* reports how many of such regions were eliminated from the sample during the application of the AGWR procedure.

In terms of overall fit, we find an improvement over the single-regime specification. Whereas the global MRW model could explain only the 18,3% (27,3% for the SEM) of overall growth variation, we find that for the identified clubs we arrive almost to explain the 80%.

The statistics in Table III suggest to choose for the OLS model the ninth (γ = 0.2; τ =0.4) or the thirteenth (γ = 0.3; τ=0.3) combination and for the SEM model the third (γ = 0.1; τ =0.3) or the eighth (γ = 0.2; τ =0.3) one.

Table III. Summary statistics for OLS and SEM models

OLS model

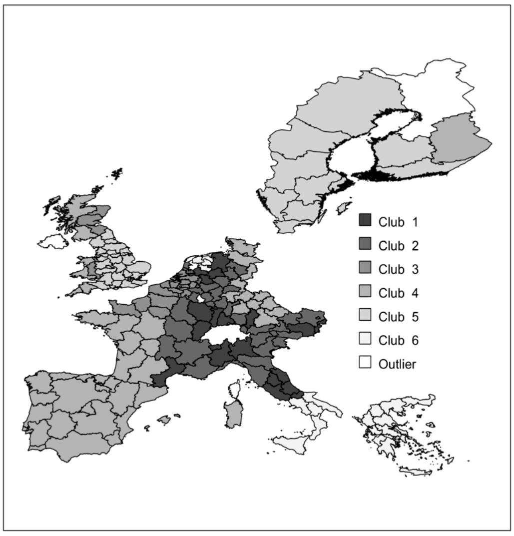
γ	τ	R^2	n. club	n. out	Max fr.	Min fr.
0.1	0.1	0.183	1	0	187	187
0.1	0.2	0.302	2	0	171	16
0.1	0.3	0.514	4	1	91	23
0.1	0.4	0.741	5	14	88	9
0.1	0.5	0.768	7	10	59	8
0.2	0.1	0.301	2	0	174	13
0.2	0.2	0.543	4	0	107	20
0.2	0.3	0.698	5	4	54	17
0.2	0.4	0.752	6	3	48	20
0.2	0.5	0.790	7	13	40	10
0.3	0.1	0.424	3	0	133	15
0.3	0.2	0.679	4	0	109	18
0.3	0.3	0.734	5	2	59	9
0.3	0.4	0.757	6	15	54	5
0.3	0.5	0.793	5	30	53	16

SEM model

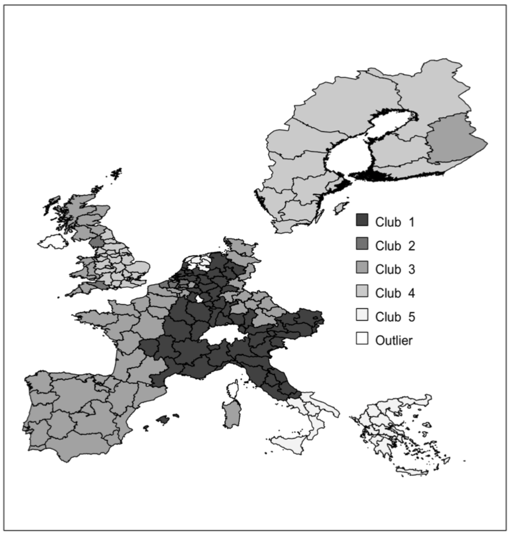
γ	τ	R^2	n. club	n. out	Max fr.	Min fr.
0.1	0.1	0.273	1	0	187	187
0.1	0.2	0.727	5	0	67	10
0.1	0.3	0.754	5	3	73	18
0.1	0.4	0.787	6	11	57	10
0.1	0.5	0.795	6	12	54	13
0.2	0.1	0.575	3	0	111	20
0.2	0.2	0.668	3	0	107	21
0.2	0.3	0.719	4	2	57	21
0.2	0.4	0.762	5	21	48	14
0.2	0.5	0.757	5	30	46	13
0.3	0.1	0.544	3	0	125	20
0.3	0.2	0.729	3	13	101	21
0.3	0.3	0.768	6	11	52	1
0.3	0.4	0.726	5	15	48	18
0.3	0.5	0.742	7	29	43	1

Figure I shows the clubs identified by AGWR for these four options: a) and b) are the clubs for the OLS models, while c) and d) are those obtained with the spatial SEM models.

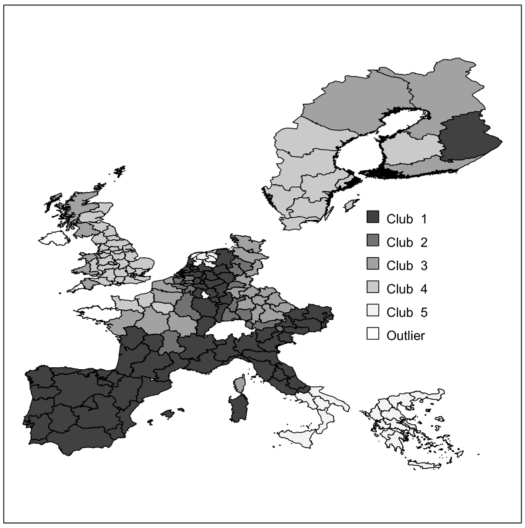
Figure I. The Clubs with OLS and SEM model



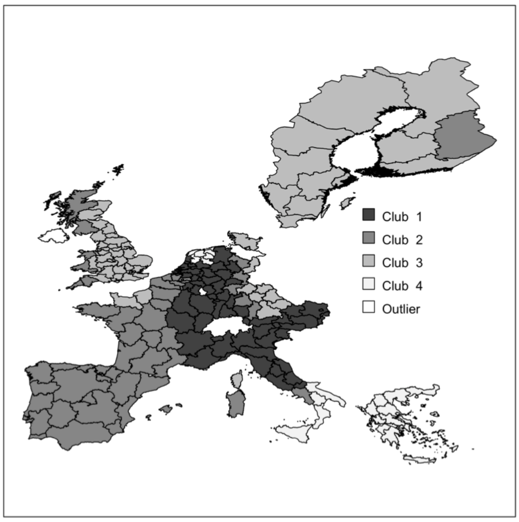
a) OLS (γ = 0.2; τ =0.4)



b) OLS (γ = 0.3; τ=0.3)



c) SEM (γ = 0.1; τ =0.3)



d) SEM(γ = 0.2; τ =0.3)

The first two and the last configurations are very similar: Club 1 in b) is partitioned into two different clubs in a); Club 2 in b) is enlarged in a) to disadvantage of Club 3 or distributed to other clubs in d).

The most striking feature of these configurations is the identification of a club of "poorest economies" composed by Southern Italy, Greece and some Eastern Netherlands regions and a club of "high literacy" regions of Scandinavian Peninsula and Central-Southern UK.

The regions identified in Figure I as outliers correspond to those excluded because unable to aggregate to other regions or to compose a club of sufficient number of observations. Table VI presents the estimates of the OLS and the SEM model for the identified clubs in b) and d) configurations. As this table indicates, there is substantial evidence that the laws of motion for growth within each subgroup are different.

Table IV. The estimated OLS and SEM models over the clubs

OLS	n	R^2	α	β	π_1	π_2	π_3	λ
Club1	59	0.434	**0.059**	**-0.012**	**-0.040**	-0.001	-0.006	1.054%
Club2	9	0.798	0.057	-0.010	-0.072	-0.003	-0.043	0.896%
Club3	58	0.196	**0.024**	**-0.003**	-0.001	-0.001	-0.004	0.290%
Club4	37	0.490	**0.028**	0.001	0.007	**0.004**	**-0.240**	0.099%
Club5	22	0.883	**0.078**	**-0.024**	-0.008	**-0.004**	0.024	1.900%

SEM	n	R^2	α	β	π_1	π_2	π_3	λ
Club1	57	0.398	**0.053**	**-0.010**	**-0.031**	-0.000	-0.000	0.896%
Club2	52	0.210	**0.025**	**-0.004**	-0.000	-0.001	0.003	0.382%
Club3	55	0.538	**0.032**	0.004	**0.043**	**0.003**	**-0.048**	0.382%
Club4	21	0.868	**0.081**	**-0.024**	-0.020	**-0.003**	**0.026**	1.896%

Note: Significant values are reported in bold.

Before interpreting our empirical results, it is important to analyze the presence of spatial dependence in the estimated models over the clubs. In fact the spatial parameter φ in the SEM club-models is never significant and the spatial tests applied on the OLS club-models reject overall the presence of spatial dependence. The identification of homogeneous groups seems to overcome the problem of spatial dependence. This result has led in fact to similar configurations and estimations independently of the starting model used in the AGWR procedure. Following these considerations we focus our comments on the results obtained for the OLS club models.

The slow convergence determined previously in the whole sample (λ = 0.654%) hid in fact different regimes speed. The estimated speed range now from almost 2% for the "poorest" club to a steady-state situation for the "high literacy" regions. Also the two larger clubs evidence vastly different estimates in the β parameter and furthermore in their speeds. Central Europe shows a speed of 1.05% with a *half-life* of 58 years, while North-West Europe has a speed of only 0.29% with a *half-life* of 231 years. The estimated coefficient β in the second club is -0.010, but insignificant. The failure to find evidence of convergence among this club can be explained by the low number of observations included in it. If we consider the corresponding enlarged Club 3 in configuration a) (Figure I), the β parameter has the same value, but become significant.

Similar heterogeneity across clubs holds for the other variables. For example, the estimated coefficient of physical capital is significant only in Club 1, while technological and demographic growth is present in Club 4 and Club 5. The coefficient of human capital is negative and significant only in Club 4, where the coefficient on population growth is positive. These two coefficient estimates for countries in Club 4 are consistent with the argument that countries whose growth had benefited from past increases in education, slow down as they reach their steady states and their education accumulation slows (Jones 2002). However the impact of human capital on economic growth remains controversial, depending on the definition and the measure of the variable, the methodology used and the time period over which the model is estimated.

Our empirical evidence suggests the presence of substantial heterogeneity of growth patterns over the identified clubs, supporting the hypothesis of convergence clubs. Moreover the estimation of heterogeneous convergence models increased significantly the goodness-of fit.

5. CONCLUSIONS

In this paper we have empirically examined the convergence hypothesis in a cross-section analysis across 187 NUTS-2 regions in 12 European countries, applying a new approach for the identification and contemporaneously estimation of the MRW model, to see whether the cross-regional growth process in Europe shows convergence clubs.

We use the adaptive geographically weighted regression methods to identify groups of countries, which obey a common model. The proposed AGWR procedure explicitly allow for cross-country parameter heterogeneity. The empirical results confirmed that such heterogeneity exists; therefore empirical analysis that fails to incorporate parameter heterogeneity can produce misleading results. In our empirical analysis we identify five clubs, with different convergence speed and different values of conditional variables.

This paper also contributes to the convergence debate by suggesting that spatial dependence can vanish or become irrelevant when parameter heterogeneity is taken into account in the model estimation. The MRW model estimated over the whole sample evidenced the presence of spatial dependence and suggested the application of a Spatial Error Model. However, when we tested the presence of spatial correlation on the heterogeneous models estimated on the clubs identified by the AGWR procedure we overall rejected such hypothesis. Therefore the usual procedure to apply spatial filter on the raw data can be a trivial technique.

REFERENCES

Anselin, L., 1988. *Spatial econometrics: Methods and Models*. Kluwer Academic Publishers: Dordrecht.

Anselin L., Bera, A.K. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DEA (eds) *Handbook of applied economic statistics*. Marcel Dekker: New York.

Anselin, L., Florax, R. 1995. *New directions in spatial econometrics*. Springer: Berlin Heidelberg New York.

Azariadis, C., Drazen, A. 1990. Threshold externalities in economic development. *Quarterly Journal of Economics* **105**: 501–526.

Barro, R.J. 1991. Economic Growth in a Cross Section of Countries, *The Quarterly Journal of Economics* **106**: 407–443.

Barro, R.J., Sala-i-Martin, X. 1995. *Economic Growth Theory*. Mc Graw-Hill: Boston.

Baumol, W. 1986. Productivity growth, convergence and welfare: what the long-run data show. *American Economic Review* **76**(5): 1072–1085.

Becker G.S., Murphy K.M., Tamura R. 1990. Human capital, fertility, and economic growth. *Journal of Political Economy* **98**(5): 12–37.

Bianchi, M. 1997. Testing for convergence: Evidence from non-parametric Multimodality Tests. *Journal of Applied Econometrics* **12**: 393–409.

Canova, F. 2004. Testing for convergence clubs in income per capita: a predictive density approach. *International Economic Review* **45**(1): 49–77.

Canova, F., Marcet, A. 1995. The Poor Stay Poor: Non-Convergence Across Countries and Regions. *CEPR working paper* 1215.

Casella, G., Berger., R.L. 2002. *Statistical Inference.* Wadsworth & Brooks (2nd ed.): Pacific Grove, California.

Caselli, F., Gerardo, E., Lefort, L. 1996. Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics. *Journal of Economic Growth* **1**: 363–389.

Cressie, N. A. C. 1993. *Statistics for spatial data*. Wiley: New York.

Desdoigts, A., 1999. Pattern of Economic Development and the Formation of Clubs. *Journal of Economic Growth* **4:** 305–330.

Durlauf, S., Johnson, P.A. 1995. Multiple Regimes and Cross-Country Behaviour. *Journal of Applied. Econometrics* **10**(4): 365–384.

Durlauf, S. 1996. On convergence and divergence of growth rates: an introduction. *Economic Journal* **106**: 1016–18.

Ertur, C., Koch, W. 2007. Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence. *Journal of Applied Econometrics* **22**: 1033–1062.

Feve, P., LePen, Y. 2000. On modeling convergence clubs. *Applied Economics Letters* **7**: 311–314.

Fotheringham, A.S., Brunsdon C., Charlton M. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley: Chichester.

Galor, O. 1996. Convergence? Inference from Theoretical Models. *Economic Journal* **106**: 1056–69.

Galor, O. 2005. From Stagnation to Growth: Unified Growth Theory. In *Handbook of Economic Growth*. : Aghion, P. and Durlauf, S.N. (eds). Elsevier: North Holland.

Galor, O. 2007. Multiple growth regimes – Insights from unified growth theory. *Journal of Macroeconomics* **29**: 470–475.

Galor, O., Zeira, J. 1993. Income Distribution and Macroeconomics. *Review of Economic Studies* **60**: 35–52.

Herault, L. 2000. Rescaled Simulated annealing–Accelerating Convergence of Simulated Annealing by Rescaling the States Energies. *Journal of Heuristics* **6**: 215–252.

Islam, N. 1995. Growth empirics: a panel data approach. *Quarterly Journal of Economics* **110**: 1127–1170.

Jones, C. I. 2002. Sources of U.S. economic growth in a world of ideas. *American Economic Review* **92**: 220–239.

Mankiw, N.G., Romer, D., Weil, D.N. 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics* **C** VII: 407–437.

Polzehl, J., Spokoiny, V. 2000. Adaptive weights smoothing with applications to image segmentation. *Journal of Royal Statistical Society* **62** (B): 335–354.

Polzehl, J., Spokoiny, V. 2003a. Image denoising: pointwise adaptive approach. *Annals of Statistics* **31**: 30–57.

Polzehl, J., Spokoiny, V. 2003b. Varying coefficient regression modeling. *Preprint 818 WIAS:* Berlin.

Postiglione, P., Benedetti, R., Lafratta G. 2010. A regression tree algorithm for the identification of convergence clubs. *Computational Statistics & Data Analysis* **54**: 2776–2785.

Quah, D. 1996. Empirics for economic growth and convergence. *European Economic Review* **40**: 1353-1375.

Quah, D., 1997. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth* **2:** 27–59.

Ramajo, J., Màrquez, M.A., Hewings, G., Salinas, M.M. 2008. Spatial heterogeneity and interregional spillovers in the European Union: Do cohesion policies encourage convergence across regions? *European Economic Review* **52:** 551–567.

Sala-i-Martin, X. 1996. The classical approach to convergence of growth analysis. *Economic Journal* **106**: 1019–1036.

Tsiddon, D. 1992. A Moral Hazard Trap of Growth. *International Economic Review* **33**: 299–321.