Productivity Polarization Across Regions in Europe

Roberto Basile

Institute for Studies and Economic Analyses

Email: r.basile@isae.it,

Address: ISAE, Piazza dell'Indipendenza, 4, 00185 – Rome,

Tel. +39-06-44482874; Fax +39-06-44482249

(May, 2007)

Abstract. The regional distribution of productivity in Europe is characterised by a Core-Periphery spatial pattern: high (low) productivity regions are in a proximate relationship with other high (low) productivity regions. Over the last twenty years, intra-distribution dynamics has generated long-run multiple equilibria with the formation of two clubs of convergence. The observed dynamics can be only marginally explained by capital accumulation and employment growth. In contrast, sectoral specialization and spatial dependence turn out to be more important factors.

Keywords: distribution dynamics, convergence, spatial dependence, specialization, Europe.
JEL classification numbers: R11, R12, C14, C21.

1. Introduction

In this work we use a continuous state-space approach to analyze the intra-distribution dynamics (thereafter, *IDD*) (Quah, 1997; Magrini, 2004; Pittau and Zelli, 2007) of regional labour productivity in Europe over the period 1980-2002. In particular, we apply a robust nonparametric conditional density estimator (Hyndman and Yao, 2002) to describe the law of motion of regional labor productivity in Europe and compute the ergodic distribution to identify long-run properties of the observed distribution dynamics (Johnson, 2004). Moreover, using a two-step approach (Lamo, 2000; Bandyopadhyay, 2003; Leonida, 2003), we try to estimate the effect of some economic determinants on the long-run distribution.

The advantages of the *IDD* approach with respect to the growth regression analysis are well known (Quah, 2006). In particular, it allows to examine directly how the whole productivity distribution changes over time and, thus, it is much more informative than the convergence empirics developed within the regression paradigm which gives information only on the dynamics of the average economy. However, some important drawbacks also characterize the literature on *IDD*. First, while the regression approach to economic convergence has been improved in many respects over the last decade[1], most of the studies based on the *IDD* approach scantly take into account the recent developments of the statistical literature on conditional density (Hyndman *et al*., 1996; Fan *et al.*, 1996; Hall *et al.*, 1999), which have highlighted the bias problems associated with the widely used standard kernel method and have proposed new estimators with

[1] Authors have proposed various estimators and econometric procedures to reduce bias and inefficiency of the regression estimates (e.g. IV and GMM estimators), to take account of spatial spill-over effects (spatial econometric techniques), to identify nonlinearities in growth behaviour (nonparametric and semiparametric regression techniques), to reduce model uncertainty (Bayesian approaches) and so on (Durlauf *et al*. 2005).

better statistical properties. In the present paper, we try to overcome this limit by using log-likelihood conditional density estimators with variable bandwidths (Hyndman and Yao, 2002)[2].

Second, while testing conditional convergence hypotheses is a very common practice within the growth regression analysis, little effort is usually devoted within the *IDD* approach to investigate the determinants of the long-run (ergodic) distribution. Quah (1997) proposed a “conditioning” scheme which allows to analyze the role of one single variable per time. More recently, some attempts to detect the joint effect of many variables appeared in the literature (Lamo, 2000; Bandyopadhyay, 2003; Leonida, 2003). In particular, a two-step approach is applied, where the first step consists of estimating a growth regression model, while in the second step the residuals from that regression are used to compute partial residuals and, thus, to estimate the conditional density functions having filtered out the effect of some growth determinants. However, all these studies use linear regression models to estimate the first step, disregarding the presence of nonlinearities in growth behavior widely highlighted in the growth regression literature (Liu and Stengos, 1999; Banerjee and Duflo, 2003; Basile, 2007b). In the present paper we overcome this limitation by applying modern semiparametric regression techniques (Wood, 2006) in order to remove the effect of growth determinants from conditional density estimations.

We propose the two above mentioned methodological improvements (robust conditional density estimators and semiparametric regressions) in order to answer many

[2] Only recently, Johnson (2005) and Fiaschi and Lavezzi (2007) have applied kernel density estimators with adaptive bandwidths to solve in some way the bias problem of the standard kernel density estimator with fixed bandwidth. Basile (2007a) compares different conditional density estimators to analyze the cross-sectional distribution dynamics of regional per-capita incomes in Europe and shows, in particular, that, while the kernel estimator with fixed bandwidth gives evidence of convergence, a modified estimator with variable bandwidth and mean-bias correction provides evidence of strong persistence and lack of cohesion.

interesting questions: Are there convergence tendencies within the group of regional economies included in the sample? If not, does one observe any specific distribution pattern? Do high-productivity regions belong to a club of high-productivity economies, while low-productivity regions languish behind (club convergence hypothesis)? Finally, what are the factors that help explain the observed dynamics of the entire distribution? In particular, does capital accumulation explain the difference between growth paths of high- and low-productivity regions? Alternatively, what is the role of industrialization? Finally, does spatial dependence matter?

The results of the analysis can be summarized as follows. First, the regional distribution of productivity in Europe is characterised by a clear Core-Periphery spatial pattern which contributes to determine a strong and increasing bimodality in the snapshot univariate density: high (low) productivity regions are in a proximate relationship with other high (low) productivity regions. Second, over the last twenty years, *IDD* has generated long-run multiple equilibria with the formation of two clubs of convergence: regions with low levels of labor productivity at the initial period have hardly managed to get close to the European average productivity in 22 years. These multiple equilibria can be only marginally explained by capital accumulation and employment growth. In contrast, sectoral specialization and spatial dependence turn out to be more important factors.

The layout of the paper is the following. In Section 2, we present a univariate analysis of regional labour productivity in Europe. In section 3, we report the results of the *IDD* analysis. In Section 4, we apply the 'multivariate' conditioning scheme and discuss the shape of the ergodic distributions computed after having removed the effects of some growth determinants. Conclusions are reported in Section 5.

2. An Exploratory Spatial Data Analysis

Most of the studies on regional convergence consider the per-capita GDP in order to measure regional unbalances. Some authors (Paci, 1997; Lopez-Bazo *et al*., 1999) have criticized this choice, observing that GDP is measured at the workplace while population at the residence and, thus, the level of per-capita GDP may lead to great distortions in some regions due to the presence of commuting patterns.[3] Based on the same considerations, here we analyse regional convergence of labour productivity, defined as the ratio between GVA (Total Gross Value Added) at constant prices 1995 and total employment for a sample of 179 NUTS-2 European regions over the period 1980-2002. Labour productivity levels are normalized with respect to the EU15 average in order to remove co-movements due to the European wide business cycle and trends in the average values. Data are drawn from the Cambridge Econometrics Dataset.

Over the period 1980-2002 the standard deviation of the relative regional labour productivity has decreased by 15%, indicating a slight σ-convergence. However, standard deviations might mask some important features of the distribution. In fact, the snapshot densities reported in Figure 1 clearly display a bimodal distribution of labour productivity both in 1980 and in 2002, indicating the existence of two clusters of regions, respectively characterised by low and high levels of labour productivity (the first mode is situated between 0.75 and 0.80 times the EU average, while the second mode is situated at around 1.10).[4] The reduction of the standard deviation is the

[3] Clear examples of this are Brussels in Belgium, Hamburg or Bremen in Germany, Inner London in the United Kingdom.
[4] Univariate densities have been estimated using the local likelihood density estimator (Loader, 1996). A variable bandwidth, selected by generalised cross validation (GCV), has been used together with a tricube kernel function. In order to allow time comparison, we have used the same span (α=0.47) for both years and evaluated the two univariate densities at the same set of data points. The evidence reported in Figure 1 is very much in line with that obtained by Fiaschi and Lavezzi (2007) using an adaptive kernel density

consequence of the decrease in the mass at the extreme tails of the distribution, while the two peaks become more pronounced in the last year.

Figure 1

Even if it is beyond the scope of this paper to identify the exact composition of the two clusters, it remains important to search for spatial patterns in the distributions of labour productivity. For this purpose, we use different measures of global and local spatial dependence as well as different mapping tools. First, Figure 2 shows a choropleth map of the percentile distribution of regional labour productivity.[5] This map allows highlighting the existence and persistence of a Core-Periphery pattern in the regional distribution of labour productivity in Europe.

Figure 2

Whether high (low) productivity regions are in a proximate relationship with other high (low) productivity regions can be more rigorously assessed by using spatial statistics. We have used distances-based binary matrices to calculate the global *G* statistic of spatial autocorrelation (Getis and Ord, 1992) defined as

$$G(d) = \sum_i \sum_j w_{ij}(d) x_i x_j \Big/ \sum_i \sum_j x_i x_j \qquad (1)$$

where x_i (x_j) is the value of labor productivity at regions *i (j)*, and w_{ij} are the elements of the binary spatial weights matrix (that is, ones for all neighbours *j* within lag distance *d* of *i* and zeros for all locations greater than *d* from *i*). A high positive value of the standardized *G* statistic, $Z(G) = \frac{G - E(G)}{V(G)^{1/2}}$, indicates that the spatial pattern is

---

estimator. Following Fiaschi and Lavezzi (2007), we have also applied a bimodality test based on the bootstrap procedure suggested by Efron and Tibshirani (1993). The p-values of this test are equal to 0.032 for the 1980 and to 0.000 for the last year, indicating the rejection of the unimodality hypothesis.

[5] In order to overcome (or, at least, to limit) the shortcomings of the crude classification of data points in few (usually 4 or 5) classes, we have imposed 100 breaks points - that is one for each percentile point - thus approximating an un-classed choropleth map (Fotheringham *et al.*, 2000).

dominated by clusters of high values, while a strong negative *Z(G)* indicates that the spatial pattern are dominated by clusters of low values.

Standardized *G* variates were computed for lag distances from 400 km (the minimum distance allowing all regions to have at least one link) up to and including 2000 km at 50 km intervals. Figure 3 shows a non-monotonic relation between distance cut-off and global spatial autocorrelation: *Z(G)* is always positive but it reaches a maximum when the cut-off distance equals 900 km; above that limit, *Z(G)* values decrease.

Figure 3

Global *G* statistic is, however, based on the assumption of stationarity or structural stability over space, which is obviously unrealistic in our context. Spatial association must be detected using local spatial autocorrelation indices which allow for local instabilities in overall spatial association. Local G_i^* indices are defined as follows (Ord and Getis, 1995):

$$G_i^*(d) = \sum_j w_{ij}(d) x_j \Big/ \sum_j x_j \quad \forall j \qquad (2)$$

with $w_{ii} \neq 0$. In our context, G_i^* is a measure of local clustering of labor productivity around region *i.* If high (low) values of *x* tend to be clustered around *i*, the standardized G_i^* will be positive (negative). No longer committed to the global pattern, local G_i^* statistics are free to characterize the spatial autocorrelation of attribute values located within a distance of each target value. Figure 4 presents standardized G_i^* variates for lag distances of 400 and 900 km for both 1980 and 2002. A typical Core-Periphery structure clearly emerges for both years: a cluster of high-productivity regions is located in the Centre of Europe (grey color), while groups of peripheral regions are characterized by negative standardized G_i^* scores (black color). Regions with a white color are those with a non-significant value of G_i^*.[6] For a cut-off distance of 900 km, the cluster of high-productivity regions is much larger, indicating that the territory becomes more homogenous.

Figure 4

Finally, standardized G_i^* scores have been used to assess whether the bimodality observed in the distribution of labour productivity is still present when we consider spatially smoothed data. Figure 5 shows snapshot densities of local G^* values. Visual inspection, supported by the bimodality test of Efron and Tibshirani (1993), suggests that the null hypothesis of unimodality cannot be accepted for spatial weights matrices based on cut-off distances lower or equal to 600 km. For higher lag distances, bimodality disappears.

Figures 5

3. Intra-distribution mobility

The univariate analysis carried out so far has allowed us to identify soma interesting features of regional labour productivity data. Nevertheless, that analysis did not give us any information on the changes of the relative position of various regions in the cross-section distribution of labor productivity over time. However, this issue is relevant for assessing the evolution of regional disparities. In order to address this drawback, it is necessary to examine the intra-distribution mobility during the study period following

[6] The critical values of the G_i^* statistic are given in Ord and Getis (1995).

the transition dynamics approach developed by Quah (1993, 1996a, 1996b, 1996c, 1997).[7]

Given the distribution of regional productivity in period t and its associated probability measure, ϕ_t, this approach consists of describing the law of motion of the stochastic process $\{\phi_t, t \geq 0\}$. If this process is assumed to be time-homogenous and first-order Markov, than the law of motion for $\{\phi_t, t \geq 0\}$ is

$$\phi_{t+\tau}(y) = \int_0^\infty f_\tau(y \mid x)\phi_t(x)dx \qquad (3)$$

where $f_\tau(y \mid x)$ is the expected density of y (the productivity levels at time $t+\tau$) conditional upon x (the productivity levels at time t). In other words, the conditional density $f_\tau(y \mid x)$ describes the probability that a given region moves to a certain state of relative productivity given that it has a certain relative productivity level in the initial period. For analyzing *IDD*, a researcher must estimate $f_\tau(y \mid x)$ and visualize the output, that is the shape of the productivity distribution at time $t+\tau$ over the range of productivity levels observed at time t.

The ergodic distribution is the limit of (3) as τ tends to infinity (Johnson, 2004):

$$\phi_\infty(y) = \int_0^\infty f_\tau(y \mid x)\phi_\infty(x)dx \qquad (4)$$

This function describes the long-term behavior of the productivity distribution: it is the density of what the cross-region productivity distribution tends towards, should the system continue along its historical path (Quah, 2006).

[7] "Distribution dynamics considers not just the time-path of income distributions – each distribution treated as a point-in-time snapshot – but also a law of motion or a mechanism for how the distribution at one time point evolves into that at a later time" (Quah, 2006, p. 14)

Operationally, the *transition dynamics approach* consists of estimating and visualizing the conditional density of y given x. The most popular approach within the *IDD* literature is the kernel density estimator with fixed bandwidths. However, this estimator has some undesirable bias properties (Hyndman *et al*., 1996) which, in the context of the *IDD* analysis, might bring to get, for example, evidence of convergence while there is persistence (Basile, 2007a). Fortunately, more robust estimators have recently been developed in the literature. In particular, Hyndman and Yao (2002) have proposed a local linear conditional density estimator which is a conditional version of Loader's (1996) density estimator used in section 2 (see Appendix 1). In the present paper we use this approach to estimate the conditional density of regional labour productivity at 2002 (the last year) given the distribution at 1980.

The results are plotted in Figure 6. These graphical methods for visualizing conditional density estimates, developed by Hyndman *et al*. (1996) and Hyndman (1996), are not common in the literature of *IDD* and, thus, a preliminary discussion on their features is necessary.[8] The first plot, called the "*stacked conditional density plot*" (figures 6A), displays a number of conditional densities plotted side by side in a perspective plot.[9] It facilitates viewing the changes in the shape of the distributions of the variable observed for the 2002 over the range of the same variable observed for the 1980. In other words, like a row of a transition matrix, each univariate density plot describes transitions over the analyzed period from a given productivity value in 1980.

[8] All of the studies on *IDD* which make use of nonparametric stochastic kernel density estimators provide three-dimensional perspective plots and/or the corresponding contour plots of the conditional density to describe the law of motion of cross-sectional distributions. In such a way, they treat the conditional density as a bivariate density function, while the latter must be interpreted as a sequence of univariate densities of relative productivity levels conditional on certain initial levels.

[9] Each univariate density plot is always non-negative and integrates to unity. Since the conditional density plot has been evaluated on an equispaced grid of 100 values over the range of x and y directions, figure 6A displays 100 stacked univariate densities.

Hyndman *et al.* (1996) note that this plot is "*much more informative than the traditional displays of three dimensional functions since it highlights the conditioning*" (p.13).

The second type of plot proposed by Hyndman *et al*. (1996) is the "*highest conditional density region*" (*HDR*) plot (figures 6B). Each vertical band represents the projection on the *xy* plan of the conditional density of *y* on *x*. In each band the 25% (the darker-shaded region), 50%, 75% and 90% (the lighter-shaded region) *HDR*s are reported. A high density region is the smallest region of the sample space containing a given probability. These regions allow a visual summary of the characteristics of a probability distribution function. In the case of unimodal distributions, the *HDR*s are exactly the usual probabilities around the mean value; however, in the case of multimodal distributions, the *HDR* displays different disjointed subregions.

The *HDR* plot is particularly suitable to analyze *IDD*. If the 45-degree diagonal crosses the 25-50% *HDR*s, it means that most of the elements in the distribution remain where they started (there is '*persistence'*). If the horizontal line traced at the one-value of the vertical axis crosses *all* the 25-50% *HDR*s, we can say that there is '*global convergence'* towards equality. If the vertical line traced at the one-value of the horizontal axis crosses *all* the 25-50% *HDR*s, we can say that there is '*global divergence'*. Finally, the presence of nonlinearities in the modal regression functions (shown in the plot as bullets) can be interpreted as an evidence in favor of the '*convergence club*' hypothesis, according to which regions catch up with one another but only within particular sub-groups.

Figure 6 shows the existence of two convergence clubs: regions sufficiently close to each other converge towards each other. The first club is composed of regions with a relative productivity level at 1980 ranging between 0.2 and 0.7 times the EU15 average;

the second club is composed of regions with a relative productivity level at 1980 ranging between 0.95 and 1.6 times the EU15 average. At the two tails of the distribution, we can observe some bimodality in the conditional density function suggesting the existence of some dualistic behaviour: a few regions with very low and very high productivity levels do not converge to any level; rather they tend to persist in their relative positions. Finally, an intermediate area, composed of regions with a relative productivity level at 1980 ranging between 0.7 and 0.95 times the EU15 average, can be classified as an area of persistence.

The shape of the ergodic distribution (Figure 7) suggests that, in the long run the European system might tend towards some reduction of regional unbalances even if the twin-peaks property remains: the first peak of the stationary distribution occurs at a slightly higher relative productivity level than that of the initial distribution; the second peak is much higher than that of both the initial and the final distributions and a decrease in the mass at the extreme tails of distribution occurs.[10]

Figure 7

4. The effect of growth determinants

4.1 The conditioning scheme

The analysis carried out so far can be interpreted as a test of the hypothesis of "*absolute convergence*", since it does not control for the heterogeneity in the structural characteristics of the regions (in terms of technologies, employment growth rates, saving rates, sectoral specialization, spatial dependence and so on; see Galor, 1996).

[10] The ergodic distribution functions reported in Figure 7 have been computed starting from the transition matrices extracted from each conditional density estimation. In order to compare univariate density functions (at 1980 and 2002) and the ergodic distribution, conditional densities have been evaluated at the same data points at which the initial density function was firstly evaluated.

Having rejected such hypothesis and having assessed that European regions tend towards different long-run equilibria (the "*club convergence*" hypothesis cannot be rejected), it remains to test the "*conditional convergence*" and the "*club conditional convergence*" hypotheses, that is it remains to understand why low-productivity regions do not tend to converge (in the long run) with high-productivity regions. In other words, our task is now to identify those factors that determine the formation of club convergence. Removing the effect of such factors, the evidence of bimodality in the ergodic distribution should disappear.

Some recent studies on *IDD* have already proposed interesting methodologies to remove the effect of some determinants of economic growth from the realized mobility dynamics across a sample of economies (Lamo, 2000; Leonida, 2003; Cheshire and Magrini, 2006). All these studies have used a two-step procedure consisting of, first, estimating a linear parametric growth regression and, then, using the residuals from this regression to simulate end-period log-relative labor productivities which, through the estimation of the conditional density function, enable to analyze the effect of different variables in shaping the dynamics of cross-regional distribution of labor productivity.

Let describe this procedure more formally. First, define $\ln y = \ln\left(GDP/EMP\right)_{t+\tau}$ and $\ln x = \ln\left(GDP/EMP\right)_t$. Thus, the growth rate of labour productivity can be expressed as $\gamma = \left(\ln y - \ln x\right)/\tau$. Now, note that the conditional density function, $f\left(y|\, x\right)$, can be written as $f\left(\exp\left(\ln x + \tau\gamma\right)|\, x\right)$. We can use this formulation to study the effect of any explanatory variable on the *IDD* of labour productivity, by defining

$$f\left(\exp\left(\ln x + \tau\left(\gamma - \hat{\gamma}\right)\right)|x\right) \qquad (5)$$

where $\hat{\gamma}$ is the growth rate predicted from a regression model.

This approach is much more appealing than the original conditioning scheme proposed by Quah (1997), since it allows to conditioning out the effect of many variables jointly. A shortcoming of this method is, however, evident: it imposes linearity in the functional form of the growth regression equation within a (flexible) nonparametric framework aimed (among other things) at identifying nonlinearities and convergence clubs. In the present paper, we propose to use, for the first step, a nonparametric or a semiparametric additive model which allows identifying nonlinearities in growth behavior.

4.2 The specification of the nonparametric growth regression additive models

The choice among which variables to include in a growth regression model varies greatly in the empirical literature. The range of potential factors suggested by economic theory is indeed very large. This raises the problem of model uncertainty that should be properly taken into account using Bayesian approaches to data analysis (Durlauf *et al*., 2005). This issue goes beyond the scope of the present paper. However, in the case of European regions, many studies have already demonstrated that sectoral composition of economic activity (Paci and Pigliaru, 1999) and spatial dependence (Lopez-Bazo *et al*. 2004; Basile, 2007b) help explain a large portion of heterogeneity in growth behaviour, together with the Solow-type conditioning factors (physical capital accumulation, employment growth and, obviously, initial conditions).[11] Therefore, we consider five different nested and non-nested models:

[11] Spatial dependence and sectoral specialization may be also important sources of multiple steady-state equilibria and club convergence. For instance, in a two-sector overlapping-generation model in which a distinction is made between consumption goods and investment goods (Galor, 1992), multiplicity of steady-state equilibria occurs under a less restrictive set of assumptions than those required in the one-sector model. Ertur and Kock (2006) propose an augmented Solow-type model with spatial externalities (spatial knowledge spillovers) between economies and provide an equation for the steady state level as well as a conditional convergence equation characterized by parameter heterogeneity: since knowledge

$$\gamma = \alpha_1 + \beta \ln x + \phi \ln\left[\frac{i_k}{n+g+\delta}\right] + \varepsilon_1 \quad (6)$$

$$\gamma = \alpha_2 + s_1(\ln x) + s_2\left(\ln\left[\frac{i_k}{n+g+\delta}\right]\right) + \varepsilon_2 \quad (7)$$

$$\gamma = \alpha_3 + s_3(\ln x, W\ln x) + s_4\left(\ln\left[\frac{i_k}{n+g+\delta}\right], W\ln\left[\frac{i_k}{n+g+\delta}\right]\right) + s_5(\rho W\gamma) + \sum_j \theta_j \ln(Y_j/Y) + \varepsilon_3 \quad (8)$$

$$\gamma = \alpha_4 + s_6(\ln x, W\ln x) + s_7\left(\ln\left[\frac{i_k}{n+g+\delta}\right], W\ln\left[\frac{i_k}{n+g+\delta}\right]\right) + s_8(\rho W\gamma) + \varepsilon_4 \quad (9)$$

$$\gamma = \alpha_5 + s_9(\ln x) + s_{10}\left(\ln\left[\frac{s_k}{n+g+\delta}\right]\right) + \sum_j \theta_j \ln(Y_j/Y) + \varepsilon_5 \quad (10)$$

The former (eq. 6) is the standard "*Solow growth regression model*". The first term on the right hand side, $\ln x$, captures the effect of interregional differences in initial aggregate productivity on interregional differences in growth rates. The second term, $\ln\left[\frac{i_k}{n+g+\delta}\right]$, captures the combined effect of the investment ratio (i_k, investment/GDP), the employment grow rate (*n*), the depreciation rate (*δ*) and the growth rate of technology (*g*). The second model specification (eq. 7) can be viewed as a "*nonlinear Solow growth regression model*", where both $\ln x$ and $\ln\left[\frac{i_k}{n+g+\delta}\right]$ enter as smooth additive terms (see, for example, Liu and Stengos, 1999).

The third specification (eq. 8) allows to test the effects of spatial dependence and sectoral specialization on regional growth and convergence. Recently, the role of spatial dependence has been theoretically and empirically investigated by Lopez-Bazo *et al*. (2004), Ertur and Kock (2007) and Basile (2007b). These studies suggest that the

---

spill-overs are local - rather than global - in scope, multiple equilibria (and, thus, convergence clubs) do emerge.

growth rate can be a negative function of initial conditions of the regions and a positive function of the initial conditions of their neighbours $(\ln x, W\ln x)$. It is also a positive function of reproducible factors accumulation rates observed within the regions and in their neighbours, $\left(\ln\left[\frac{i_k}{n+g+\delta}\right], W\ln\left[\frac{i_k}{n+g+\delta}\right]\right)$. As in Basile (2007b), the effect of these variables are captured by introducing nonparametric interaction terms in the model. Compared to eq. 7, eq. 8 includes another spatially lagged term, $s_5(W\gamma_y)$, which represents the rate of growth in the neighbouring regions.

Eq. 7 includes some linear terms, $\sum_j \ln(Y_j/Y)$, where Y_j is the value added produced in sector *j* by each region. These terms should allow us to estimate the linear effect of differences in the sectoral composition of GDP on regional variations of productivity growth rates. In particular, this model allows us to test the hypothesis that if the productivity growth rate of a particular sector lags that of the other sectors, regions with largest specialization in that sectors tend to exhibit the slowest aggregate productivity growth. Those scholars that have previously tested such hypothesis (Dowrick and Gemmel, 1991; Leonida, 2003) have focused on broadly-defined sectors, such as agriculture and industry. Here, we try to use more finely defined sector specialization exploiting Cambridge Econometrics information on sectoral value-added of European regions in 14 sectors[12]. Of course, it is not convenient to introduce so many variables in our eq. 8 because of the potentially high multiple correlation between these

[12] The sectors considered in the dataset are: Agriculture, Forestry and Fishing (categories A+B of the NACE rev. 1, the Classification of Economic Activities in the European Community); Mining and Energy (C); Food, Beverages and Tobacco (DA); Textile and Clothing (DB+DC); Electronics (DL); Fuels, Chemicals, Rubber and Plastic products (DF+DG+DH); Transport equipment (DM); Other manufacturing (DD+DE+DI+DJ+DK+DN); Construction (F); Wholesale and Retail (G); Hotels and Restaurants (H); Transport and Communications (I); Financial Services (J); Other market services (K); Non-market services (L+M+N+O).

variables. Thus, we have used a principle component analysis in order to reduce the number of dimensions without loss of information.[13] Precisely, the first 5 components explain 73% of the overall variability. In a preliminary analysis, we introduced these 5 components in the regression model and found that only the first 3 were significant. These components (which explain 67% of the overall variability) have also a quite clear interpretation. The first one (*PC-1*) indicates a very strong dependence between electronics, transport and equipment and other manufacturing (a miscellanea of industries including metal, paper, wood, non-metallic mineral products, and machinery); thus, we define this components as "*Specialization in high-tech and scale-intensive sectors*". The second component (*PC-2*) indicates a strong dependence between agriculture, hotel and restaurant, textile, clothing and footwear and, thus, we define it as "*Specialization in low-tech and traditional sectors*". The third component identifies strong dependence between service sectors and it is defined as "*Specialization in market and non market services*". This empirical strategy allows us to overcome the long-debated issue on whether industrialized economies outperform rural economies. The role of specialization within the manufacturing on growth behavior, and in particular the effect of specialization in high-tech vs. low-tech or traditional sectors, represents a more important issue to be investigated. Finally, model defined in equations 9 and 10 are nested models into eq. 8.

Additive models specified in eq. 7-10 have been estimated using a mixture of parametric linear terms and bivariate thin-plate regression splines and applying the method described in Wood (2006) that allows integrated smoothing parameter selection

[13] We might have introduced the sector composition variables as smooth terms. However, again, estimating 14 smooth additional terms would have required a large amount of degree of freedom. Moreover, we cannot exclude the existence of "concurvity" (the nonparametric analogue of collinearity), which may lead to statistically unstable contributions of variables to additive models and, thus, may impact the interpretation of the additive fits.

via *GCV* (see Appendix 2). This method (implemented in the R package *mgcv*) helps overcome the difficulties of model selection typical of the AM framework based on back-fitting developed by Hastie and Tibshirani (1990).

4.3 Regression results

Tables 1 and 2 report the results and a series of diagnostic statistics for the five models. The proportion of deviance explained ranges from 45.1% (linear Solow model) to 74.7% (model 3 - eq. 8), while the *GCV* score reaches the lowest level with model 3, clearly suggesting that the most general specification encompasses all the others. Moreover, the hypotheses of normality, constant variance and no spatial dependence in the residuals cannot be rejected only in the cases of models 3 and 4; the residuals from models 2 and 5 are strongly nonlinear and show significant spatial dependence, while those from model 1 have also non-constant variance. All this suggests that spatial dependence must be explicitly taken into account in order to avoid misspecification problems and that sectoral specialization partially contributes to explain heterogeneity in regional growth behavior in Europe.

Tables 1 and 2

The linear coefficients of model 1 (the standard Solow growth model) have the expected sign and are significantly different from zero. The coefficient β associated to the linear term $\ln x$ is –0.987, while the coefficient ϕ connected to the linear term $\ln\left[\frac{s_k}{n+g+\delta}\right]$ is 0.311. The *F*-tests for the overall significance of the smoothed terms in AMs 2-5 have p-values lower than 0.001, while the number of effective degrees of freedom (e.d.f.) suggests that the relationship between growth rates and Solow-growth determinants is far from being linear. Finally, the parameters associated to the linear terms of the semiparametric model 3 are respectively 0.071, -0.057 and 0.137 and are all significantly different from zero. Thus, regions with a higher specialization in "*high-tech and scale-intensive sectors*" or in "*market and non-market services*" have expected higher growth rates, while regions with a higher specialization in "*low-tech and traditional sectors*" have lower expected growth rates.

To save space, we only discuss graphical results for model 3 (Figure 8). The vertical axis reports the scale of productivity growth rates; the axes on the plane report the scale of each independent variable and of its correspondent spatial lag. Figure 8A shows the estimated effect of the interaction between $\ln x$ and the correspondent spatial lag, $W \ln x$, on the growth of labour productivity. It clearly suggests that regions surrounded by higher productivity regions have higher expected growth rates than regions with a lower-productivity neighbourhood. Thus, while very low-productivity regions have generally higher expected growth rates, as it is predicted by the Solow growth model, those with high-productivity neighbours have the highest rates of growth. Moreover, even very high-productivity regions (which are closer to their steady state and, thus, have lower margins for catching up) have chance to grow faster when surrounded by high-productivity regions.

As expected, the effect of the interaction between $\ln x$ and $W \ln x$ is also characterized by strong nonlinearities. However, one can also observe that the marginal effect of $\ln x$ appears to be quasi-linear, in contrast with the two-convergence club picture depicted in Figure 6. But, it is important to say that this result is obtained only after having controlled for the effect of spatial dependence. In fact, the marginal effect of $\ln x$ from model 5 (Figure 9) allows to identify two negatively-sloped segments, indicating two groups of regions converging towards different steady states, and a zerosloped segment in the middle, indicating the presence of a non-converging group of middle-productivity regions.

Figure 8B shows the marginal effect of the interaction between $\ln\left[\frac{i_k}{n+g+\delta}\right]$ and $W\ln\left[\frac{i_k}{n+g+\delta}\right]$. First, some nonlinearities in the effect of the rate of capital accumulation are clearly detected: an increase in the rate of capital accumulation is associated with an increase in growth rate only when $\ln\left[\frac{i_k}{n+g+\delta}\right]$ is above a certain threshold. Moreover, the growth rate of a region is also a positive function of the capital accumulation rate in the neighbours.

Model 3 included another term, $s_5\left(W\gamma_y\right)$, measuring the smooth effect of the rate of growth in the neighbouring regions, the so-called spatial externalities effect. Of course, this term cannot be considered as exogenous. Thus, miming the spatial two-stage least-square procedure (Kelejian and Prucha, 1998), we have adopted an instrumental variable approach, using smoothed spatially lagged exogenous variables as instruments for the spatially lagged dependent variable.[14] Figure 8C shows the fitted smooth functions $\hat{s}_5\left(W\gamma_y\right)$ alongside Bayesian confidence intervals (Wood, 2004). The vertical axis reports the scale of the expected values of relative regional growth rate; the horizontal axis reports the scale of the spatial lag of the relative growth rate. The *F* test suggests that overall this term has a significant nonlinear effect on the expected growth rate. In particular, the spatial autocorrelation curve first increases and then it slopes

[14] Specifically, we have used smooth terms of the second order spatial lags of $\ln x$ and of $\ln\left[\frac{i_k}{n+g+\delta}\right]$, as well as first order spatial lags of *PC-1*, *PC-2* and *PC-3*.

downwards (Figure 8C). Thus, it seems that positive spatial spillovers occur up to a certain threshold, beyond which competition effects prevail so that regions surrounded by very-high-growth regions have a lower expected growth rate than regions surrounded by medium-growth regions.

4.4 Conditioned ergodic distributions

The econometric results discussed in the previous section have provided strong evidence of nonlinearities in the effect of initial conditions and of capital accumulation, thus suggesting that the linear Solow growth model suffers from misspecification problems. They have also highlighted the importance of sector specialization, casting some doubt on the validity of the one-sector growth model. Moreover, some specifications have allowed identifying the effect of interactions between the characteristics (initial conditions and physical capital investment) of each region and those of its neighbours, confirming the prediction of recently developed spatial neoclassical growth models (Ertur and Kock, 2007).

This section reports the results of the ergodic distributions computed after having removed the effect of growth determinants. In practice, we have first re-estimated eq. 7-10 without the smooth term $s(\ln x)$ and the constant term in order to compute the prediction $\hat{\gamma}$ and, thus, estimate conditional densities as in eq. 5. Then, we have estimated 'conditioned' ergodic distributions using the transition matrices extracted from each conditional density estimation and compared them with the 'unconditioned' ergodic distribution (Figure 7).

Figure 7B reports the ergodic distribution computed after having removed the smooth effect of capital accumulation from the actual productivity growth rate. This graph suggests that, even if capital accumulation has a positive and significant effect on productivity growth, it does not contribute to explain the long-run distribution of regional labor productivity: the shape of the ergodic distribution remains bimodal with the two peaks almost at the same point and at the same height as in Figure 7A.

Some differences between the ‘unconditioned’ and ‘conditioned’ ergodic distributions emerge when the effect of sector specialization is removed from the actual productivity growth rate (Figure 7E). In particular, even if the ergodic distribution remains clearly bimodal, the first peak becomes less pronounced while the second one moves to the right, indicating some more regional convergence in the long run.

A clear transformation of the ergodic distribution appears only after having filtered out the effect of spatial dependence. Figures 7C and 7D show how the evidence of convergence clubs (bimodality) disappears when we control for spatial interaction in empirical growth regression (eq. 8-9), even if residual bumps characterize the long-run distribution.

5. Discussion and conclusions

In this paper we have used a continuous state-space approach to analyze the distribution dynamics of regional labor productivity in Europe over the period 1980-2002. The results confirm the existence of multiple equilibria in regional growth behavior in Europe with the formation of two clubs of convergence, which have also a clear spatial patterns: high productivity regions mainly located in the core of Europe tend to converge to high productivity levels, while most of the peripheral regions belong to the low productivity club of convergence.

Using a two-step approach, we have also analyzed the determinants of the shape of the long-run (ergodic) distribution of regional labor productivity. The results suggest that observed dynamics can be only marginally explained by capital accumulation. In contrast, spatial dependence is primarily responsible for the bimodality in the long-run distribution of labor productivity. Finally, our findings imply that sector specialization matters, thus casting some doubt on the validity of the one-sector neoclassical growth model.

Policy implications

Appendix 1: Local linear conditional density estimators

The most common estimator of the conditional density widely used in the literature of *IDD* is the kernel estimator, firstly proposed by Rosenblatt (1969). Recently, Hyndman *et al.* (1996) have explored its properties. They define:

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

where

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

Thus, the conditional density estimator can be interpreted as the Nadaraya-Watson kernel regression of $K\left(\frac{\|y - Y_i\|_y}{b}\right)$ on X_i. As it is well known, the Nadaraya-Watson estimator can present a large bias both on the boundary of the predictor space, due to the asymmetry of the kernel neighbourhood, and in its interior, if the true function has substantial curvature or if the design points are very irregularly spaced.

Given the undesirable bias properties of the kernel smoother, Hyndman and Yao (2002) have proposed a local likelihood conditional density estimator, which is a conditional version of Loader's (1996) density estimator. Let

$$R(\beta_0, \beta_1; x, y) = \sum_{i=1}^{n} \left\{ K\left(\frac{\|y - Y_i\|_y}{b}\right) - \exp\left(\beta_0 - \beta_1 (X_i - x)\right) \right\}^2 K\left(\frac{\|x - X_i\|_x}{a}\right)$$

where $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1\right)$ is that value of β which minimizes $R(\beta_0, \beta_1; x, y)$. The local linear density estimator at a focal point x is then defined as $\hat{f}(y \mid x) = \hat{\beta}_0$. This estimator has a smaller bias than the Nadaraya-Watson estimator. All conditional densities in the present paper have been estimated using Hyndman and Yao (2002). Optimal bandwidths, a and b, have been selected using the method developed by Bashtannyk and Hyndman (2001) based on GCV.

Appendix 2: Additive models with integrated model selection

Additive models (AM) provide a framework for nonparametric and semiparametric modeling. In general the model has a structure something like:

$$y_i = \mathbf{X}_i^* \theta + s_1(x_{1i}) + s_2(x_{2i}) + s_3(x_{3i}, x_{4i}) + \ldots + \varepsilon_i \qquad \varepsilon_i \sim i.i.d.\ N\left(0, \sigma^2\right) \qquad \text{(A1)}$$

where y_i is an univariate response continuous variable, $\mathbf{X}_i^*$ is a vector of strictly parametric components, $\boldsymbol{\theta}$ is the corresponding parameter vector and $s_j(..)$ are smooth functions of the covariates, x_j. The estimated function $\hat{s}(..)$ can reveal possible nonlinearities in the effect of x_j.

The most popular approach for estimating AM is the back-fitting algorithm proposed by Hastie and Tibshirani (1990). This approach, however, presents some shortcomings with respect to the issues of model selection and inference. Wood (2000, 2006) and Wood and Augustin (2002) have recently proposed a new method to estimate AM with spline based penalized regression smoothers which allows for automatic and integrated smoothing parameters selection via Generalized Cross Validation (GCV). Wood has implemented this approach in the R package *mgcv*.

In the case of a model containing one smooth function of one covariate ($y_i = s(x_i) + \varepsilon_i$), the penalized regression spline arises as the solution to the following optimization problem:

$$\min \left\| \mathbf{X}\boldsymbol{\beta} - \mathbf{y} \right\|^2 + \lambda \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta} \qquad \text{(A2)}$$

w.r.t. $\boldsymbol{\beta}$ (the parameter vector). $\|\cdot\|$ is the Euclidean norm and $\mathbf{S}$ is a positive semi-definite matrix depending on the basis functions evaluated at x. Given λ (a constant smoothing parameter), the solution to (A2) is:

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}^T \mathbf{X} + \lambda \mathbf{S} \right]^{-1} \mathbf{X}^T \mathbf{y} \qquad \text{(A3)}$$

A crucial issue in the use of smoothing splines is the selection of parameter λ, controlling the trade-off between fidelity to the data and smoothness of the fitted spline. Generalized Cross Validation (GCV) is the most common method used to choose the smoothing parameter:

$$GCV(\lambda) = \frac{n \left\| \mathbf{y} - \mathbf{A}\mathbf{y} \right\|^2}{\left[n - tr(\mathbf{A}) \right]^2} \qquad \text{(A4)}$$

where $\mathbf{A}$ is the hat matrix for the model being fitted: $\mathbf{X}\left(\mathbf{X}^T \mathbf{X} + \sum_i \lambda \mathbf{S}_i \right)^{-1} \mathbf{X}^T$, and the term tr($\mathbf{A}$) gives the estimated degrees of freedom of the model. The best λ is the one that minimizes GCV.

When there are two or more smoothed terms (e.g. $y_i = s_1(x_{1i}) + s_2(x_{2i}) + \varepsilon_i$), the selection of the smoothing parameters becomes less straightforward. Consider first the back-fitting algorithm proposed by Hastie and Tibshirani (1990). It consists of estimating each term by iteratively smoothing partial residuals from the AM w.r.t the covariate that the smooth relates to. Thus, given the bandwidth of the smoothers, the estimation of smooth terms becomes straightforward with back-fitting. However, “estimation of that bandwidth is hard to integrate into a back-fitting approach” and the choice of the degree of smoothness of each term in the model becomes arbitrary (Wood and Augustin, 2002, p. 2). To overcome this problem, Wood (2000) provides a methodology to choose automatically multiple smoothing parameters by GCV, as in the single-penalty case. First, he suggests to write the model fitting problem with an extra “overall” smoothing parameter (ρ) controlling the trade-off between model fit and overall smoothness:

$$\min \left\| \left(\mathbf{X}\boldsymbol{\beta} - \mathbf{y} \right) \right\|^2 \rho + \sum_{j=1}^{m} \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta} \qquad \text{(A5)}$$

w.r.t. $\boldsymbol{\beta}$ subject to the linear constraint $\mathbf{C}\boldsymbol{\beta} = 0$, where $\mathbf{C}$ is a matrix of known coefficients defining the constraints. The smoothing parameters, ρ and λ, are estimated by minimizing the GCV score:

$$GCV\left(\rho, \boldsymbol{\lambda}\right) = \frac{n \left\| \left(\mathbf{y} - \mathbf{A}\left(\rho, \boldsymbol{\lambda}\right) \mathbf{y} \right) \right\|^2}{\left[n - tr\left(\mathbf{A}\left(\rho, \boldsymbol{\lambda}\right) \right) \right]^2} \qquad \text{(A6)}$$

w.r.t. the relative smoothing parameters, λ/ρ. Problems (A4) and (A6) are solved iteratively:

1. given the current estimates of the relative smoothing parameters, estimate the overall smoothing parameter using single smoothing parameter methods;
2. given the overall smoothing parameter, update the logarithms of the relative smoothing parameters simultaneously using Newton’s method.

Therefore, with this method, the smoothing parameters for each smooth term in the model are chosen simultaneously and automatically as part of model fitting by minimizing the GCV score of the whole model.

So far, the approach for estimating an AM with the automatic model selection developed by Wood (2000) has been described for the simple case of one dimensional basis functions. Wood and Augustin (2002) and Wood (2003) have extended this approach to the cases of multi-dimensional bases, in particular to the thin plate regression splines and to the tensor products. Specifically, Wood (2006) recommends to use thin-plate regression splines for smooth interactions of quantities measured in the same units, while he suggests to use tensor products for smooth interactions of quantities measured in different units, or when very different degrees of smoothing are appropriate relative to different covariates.

Bibliography

Azariadis C, Drazen A. 1990. Threshold externalities in economic development. *Quaterly*

Journal of Economics **105**: 501-526.

Basile R., 2007a, Intra-distribution dynamics of regional per-capita income in Europe: evidence from alternative conditional density estimators, ISAE wp n. 75.

Basile R., 2007b, Regional Economic Growth in Europe: a Semiparametric Spatial Dependence Approach, mimeo.

Bandyopadhyay S., 2003, Polarisation, Stratification and Convergence Clubs: Some Dynamics and Explanations of Unequal Economic Growth across Indian states, mimeo.

Banerjee, A. and E. Duflo (2003), 'Inequality and growth: what can the data say?', *Journal of Economic Growth*, 8(3), 267-300.

Stefano Magrini & Paul Cheshire, 2006. "European Urban Growth: now for some problems of spaceless and weightless econometrics," Working Papers 23_06, University of Venice "Ca' Foscari", Department of Economics

Dowrick, S. and N. Gemmell, (1991), "Industrialisation, Cathing up and Economic Growth: A Comparative Study Across theWorld's Capitalist Economies", The Economic Journal, 101.

Durlauf SN, Johnson PA, Temple JRW. 2005. Growth Econometrics. In *Handbook of Economic Growth*, Volume 1A, Aghion P, Durlauf SN (eds). North-Holland: Amsterdam.

Efron B. and R. Tibshirani (1993), An introduction to the bootstrap. London: Chapman and Hall.

Ertur C. and Koch W., "Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence", *Journal of Applied Econometrics*, forthcoming.

Fan J, Yao Q, Tong H. 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**: 189–206.

Fiaschi D. and Lavezzi M. (2007), Productivity Polarization and Sectoral Dynamics in European Regions, mimeo.

Fotheringham, A.S., Brunsdon, C. and Charlton M. (2000), Quantitative Geography: Perspectives on Spatial Data, Sage Publications Ltd.

Galor O. 1996. Convergence? Inferences from theoretical models. *Economic Journal* **106**: 1056-1069

Getis. A, Ord, J. K. 1992 The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24, 195-.

Getis, A. and Ord, J. K. 1996 Local spatial statistics: an overview. In P. Longley and M. Batty (eds.) *Spatial analysis: modelling in a GIS environment* (Cambridge: Geoinformation International), 261–277.

Hall P, Wolff R, Yao Q. 1999. Methods for estimating a conditional distribution function. *Journal of American Statistical Association* **94**: 154–163.

Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*, New York: Chapman and Hall

Hyndman RJ. 1996. Computing and Graphing Highest Density Regions. *The American Statistician* **50**: 120-126.

Hyndman RJ, Bashtannyk DM, Grunwald GK. 1996. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics* **5:** 315-336.

Hyndman RJ, Yao Q. 2002. Nonparametric estimation and symmetry tests for conditional density functions. *Journal of Nonparametric Statistics* **14**: 259-278

Lamo A. 2000. On Convergence Empirics: Some Evidence for Spanish Regions. *Investigaciones Economicas* **24**: 681-707.

Leonida, L., 2003, On The Effects of Industrialization on Growth and Convergence Dynamics in Italy (1960-95), mimeo.

Liu, Z. and T. Stengos (1999), 'Non-Linearities in Cross-Country Growth Regressions: a Semiparametric Approach', *Journal of Applied Econometrics*, 14, 527-538.

Loader CR. 1996. Local likelihood density estimation. *The Annals of Statistics* **24**: 1602–1618.

Lòpez-Bazo, E., E.Vayà and M. Artìs (2004), 'Regional externalities and growth: evidence from European regions', *Journal of Regional Science*, 44(1), 43-73.

Magrini S. 2004. Regional (Di)Convergence. In *Handbook of Regional and Urban Economics,* Henderson V, Thisse JF (eds.). North-Holland: Amsterdam.

Ord, J. K. and Getis, A. 1995 Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, 27, 286–306.

Paci, R., (1997), "More similar and less equal. Economic growth in the European regions", Weltwirtschaftliches Archiv, 4.

Paci, R. and F. Pigliaru (1999), 'European Regional Growth: Do Sectors Matter?' in Adams, J. and F. Pigliaru (ed.), *Economic Growth and Change. National and Regional Patterns of Convergence and Divergence*, Cheltenham, UK: Edward Elgar.

Pittau MG, Zelli R. 2006. Income dynamics across EU regions: empirical evidence from kernel estimator. *Journal of Applied Econometrics* **21**: 605-628.

Quah D. 1993. Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics* **95**: 427–443.

Quah D. 1996a. Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Economic Journal* **106**: 1045-55.

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

Quah D. 1996c. Convergence Emprics Across Economies with (Some) capital Mobility.

Journal of Economic Growth **1**: 95-124.

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

Quah D. 2006. Growth and distribution. *Mimeo LSE Economc department* August.

Unwin, A. and Unwin, D. (1998) Spatial Data Analysis with Local Statistic Journal of the Royal Statistical Society: Series D (The Statistician) 47 (3).

Johnson PA., 2004. *A continuous state space approach to "convergence by parts".* Department of Economics, Vassar College, Poughkeepsie, NY.

Wood, S.N. (2003) Thin plate regression splines. J.R.Statist.Soc.B 65(1):95-114.

Wood, S.N. (2006), *Generalized Additive Models. An Introduction with R*, Boca Ratom: Chapman & Hall/CRC.

Table 1 – Results of additive models

	Linear and smoothed terms	Coefficients (p-values)	F test (p-values)	e.d.f.	Deviance	AIC	GCV
1	$\ln x$	-0.987 (0.000)			45.1	-1,425	0.0201
	$\ln\left[\frac{s_k}{n+g+\delta}\right]$	0.311 (0.045)					
2	$s_1(\ln x)$		12.6 (0.000)	3.9	54.7	-1,443	0.0183
	$s_2\left(\ln\left[\frac{s_k}{n+g+\delta}\right]\right)$		3.0 (0.002)	6.1			
3	$s_3(\ln x, W\ln x)$		6.7 (0.000)	10.2	74.7	-1,509	0.0130
	$s_4\left(\ln\left[\frac{s_k}{n+g+\delta}\right], W\ln\left[\frac{s_k}{n+g+\delta}\right]\right)$		2.9 (0.000)	11.4			
	$s_5(\rho W\gamma)$		3.2 (0.002)	4.4			
	$\ln(Y_1/Y)$	0.068 (0.000)					
	$\ln(Y_2/Y)$	-0.052 (0.019)					
	$\ln(Y_3/Y)$	0.078 (0.004)					

continue …

Table 1 – Results of additive models

	Linear and smoothed terms	Coefficients (p-values)	F test (p-values)	e.d.f.	Deviance	AIC	GCV
4	$s_6(\ln x, W \ln x)$		13.7 (0.000)	2.6	67.8	-1,485	0.0146
	$s_7\left(\ln\left[\frac{s_k}{n+g+\delta}\right], W \ln\left[\frac{s_k}{n+g+\delta}\right]\right)$		2.3 (0.001)	12.1			
	$s_8(\rho W \gamma)$		5.8 (0.000)	4.7			
5	$s_9(\ln x)$		14.5 (0.000)	5.8	62.8	-1,471	0.0157
	$s_{10}\left(\ln\left[\frac{s_k}{n+g+\delta}\right]\right)$		3.9 (0.000)	5.0			
	$\ln(Y_1/Y)$	0.072 (0.000)					
	$\ln(Y_2/Y)$	-0.087 (0.000)					
	$\ln(Y_3/Y)$	0.086 (0.001)					

Notes:

- **Model 1**: Linear Solow model; **Model 2**. Nonlinear Solow model; **Model 3**: Spatial nonlinear Solow model augmented with sectoral specialization; **Model 4**: Spatial nonlinear Solow model; **Model 5**. Nonlinear Solow model augmented with sectoral specialization.
- **Coefficients** refer to parametric terms.
- **F tests** are used to investigate the overall ('approximate') significance of smooth terms.
- **E.d.f.** (effective degrees of freedom) reflect the flexibility of the model. An e.d.f. equals to 1 suggests that the smooth term can be approximated by a linear term.
- **Deviance** is the proportion of deviance explained.
- **AIC** is the ('approximate') Akaike Information Criterion.
- The **GCV score** (x 1000) provides a criterion for choosing the model specification among several different possible alternatives. Thus, the decision to remove or maintain a term is based on comparison of GCV scores and the model which minimizes the GCV is preferred.

Table 2 - Diagnostics of residuals

		Model 1	Model 2	Model 3	Model 4	Model 5
Normality	JB	22.860 (0.000)	31.674 (0.000)	1.196 (0.549)	2.407 (0.300)	22.129 (0.000)
Spatial depend.	MC-I	400 km: 0.001 900 km: 0.002	400 km: 0.001 900 km: 0.004	400 km: 0.118 900 km: 0.361	400 km: 0.593 900 km: 0.607	400 km: 0.002 900 km: 0.024
	GC-R	400 km: 0.002 900 km: 0.000	400 km: 0.001 900 km: 0.000	400 km: 0.108 900 km: 0.333	400 km: 0.570 900 km: 0.641	400 km: 0.002 900 km: 0.017
Constant variance	F test (p-values)	3.636 (0.000)	0.317 (0.574)	1.222 (0.289)	1.935 (0.058)	1.366 (0.207)

Notes:

- **Model 1**: Linear Solow model; **Model 2**. Nonlinear Solow model; **Model 3**: Spatial nonlinear Solow model augmented with sectoral specialization; **Model 4**: Spatial nonlinear Solow model; **Model 5**. Nonlinear Solow model augmented with sectoral specialization.
- The **normality test** is based on Jarque-Bera (JB) statistics (p-value in parenthesis).
- The tests of **spatial dependence** (using two different distance neighbors weights matrices) are based on a Monte Carlo Simulation of Moran's I (MC-I) and on Geary C test under randomization (C-R) (p-values).
- The **test of constant variance** of the residuals is based on the estimation of the simple model $|\hat{e}| = \alpha + s(\hat{y}) + \varepsilon$, where $|\hat{e}|$ is the absolute value of the residuals of the model and $\hat{y}$ is the vector of fitted values. Under the null hypothesis of constant variance, the smooth term $s(\hat{y})$ must be estimated with one degree of freedom and, according to a F test, should not have a significant effect on $|\hat{e}|$.

Figure 1 – Univariate density

Local likelihood density estimation with variable bandwidth (Loader, 1996)

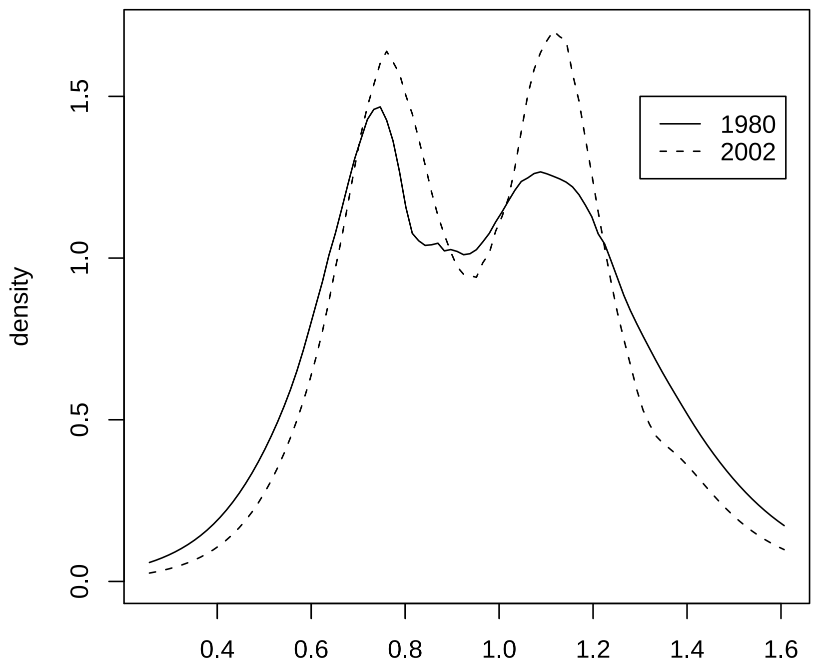
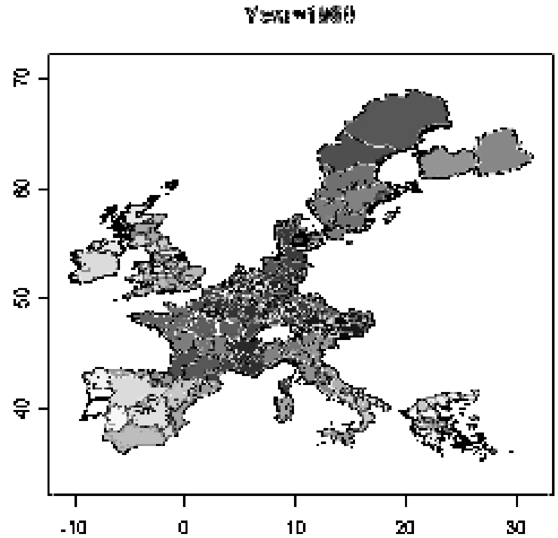
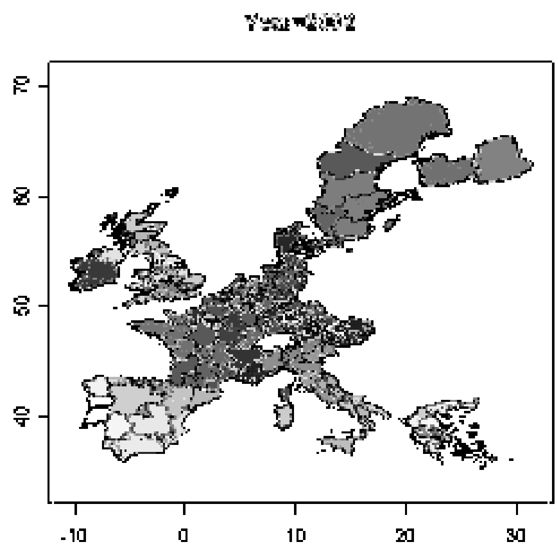


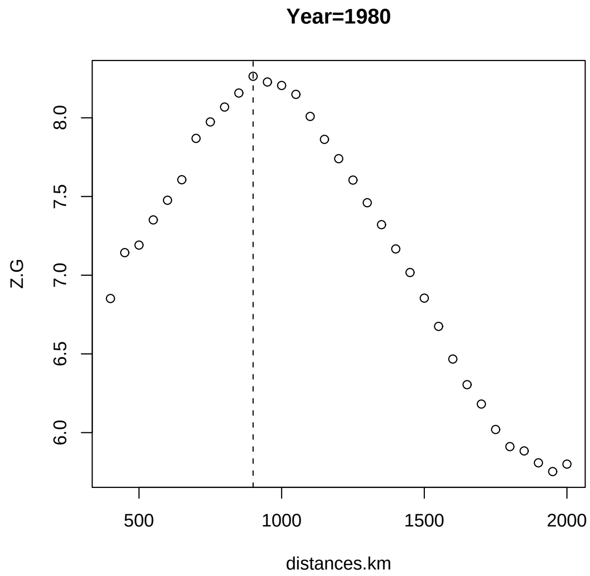
Figure 2 – Choropleth maps of the distribution of labour productivity in 1980 and 2002





Note: regional productivity levels have been classified using 100 break-points. The intensity of the grey color varies with the variable of interest.

Figure 3 – Global G statistics



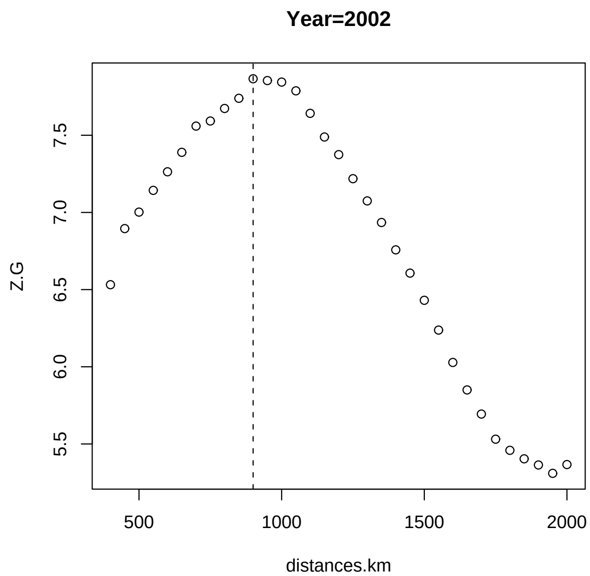


Figure 4 – Maps of G* indices

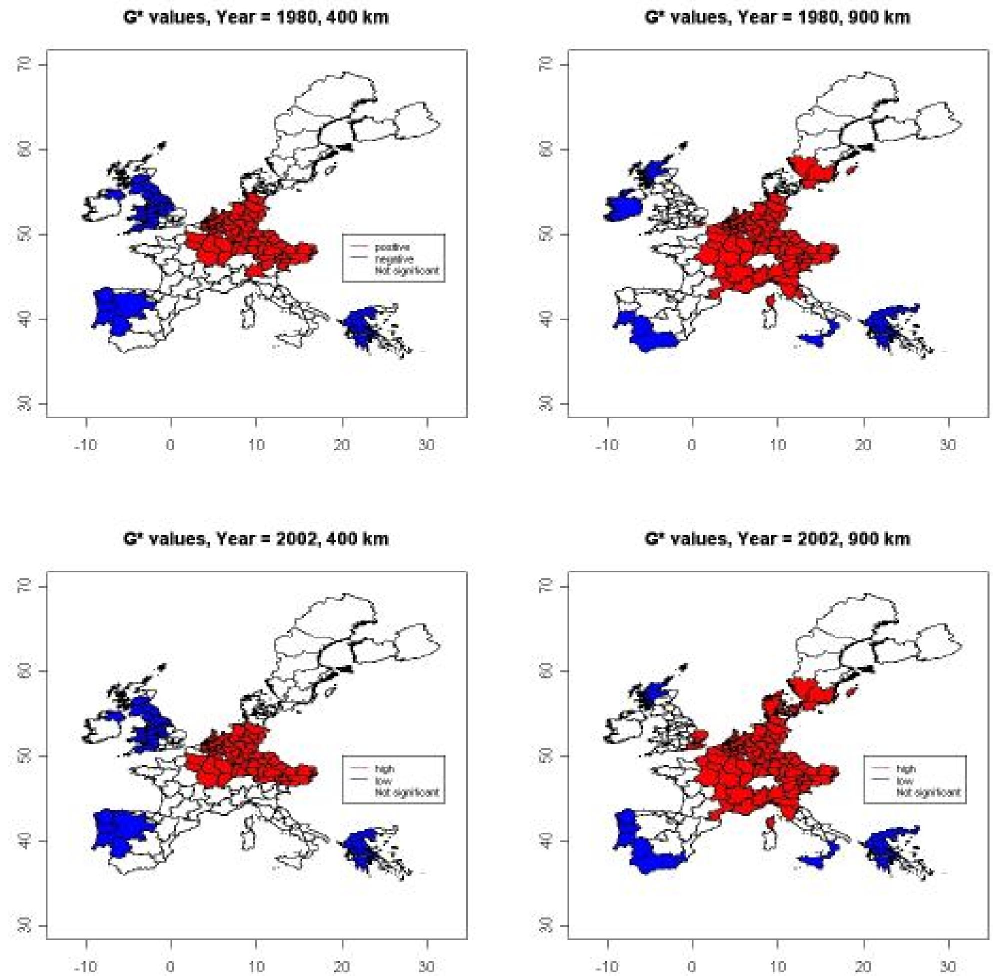


Figure 5 – Univariate densities of local G* indices

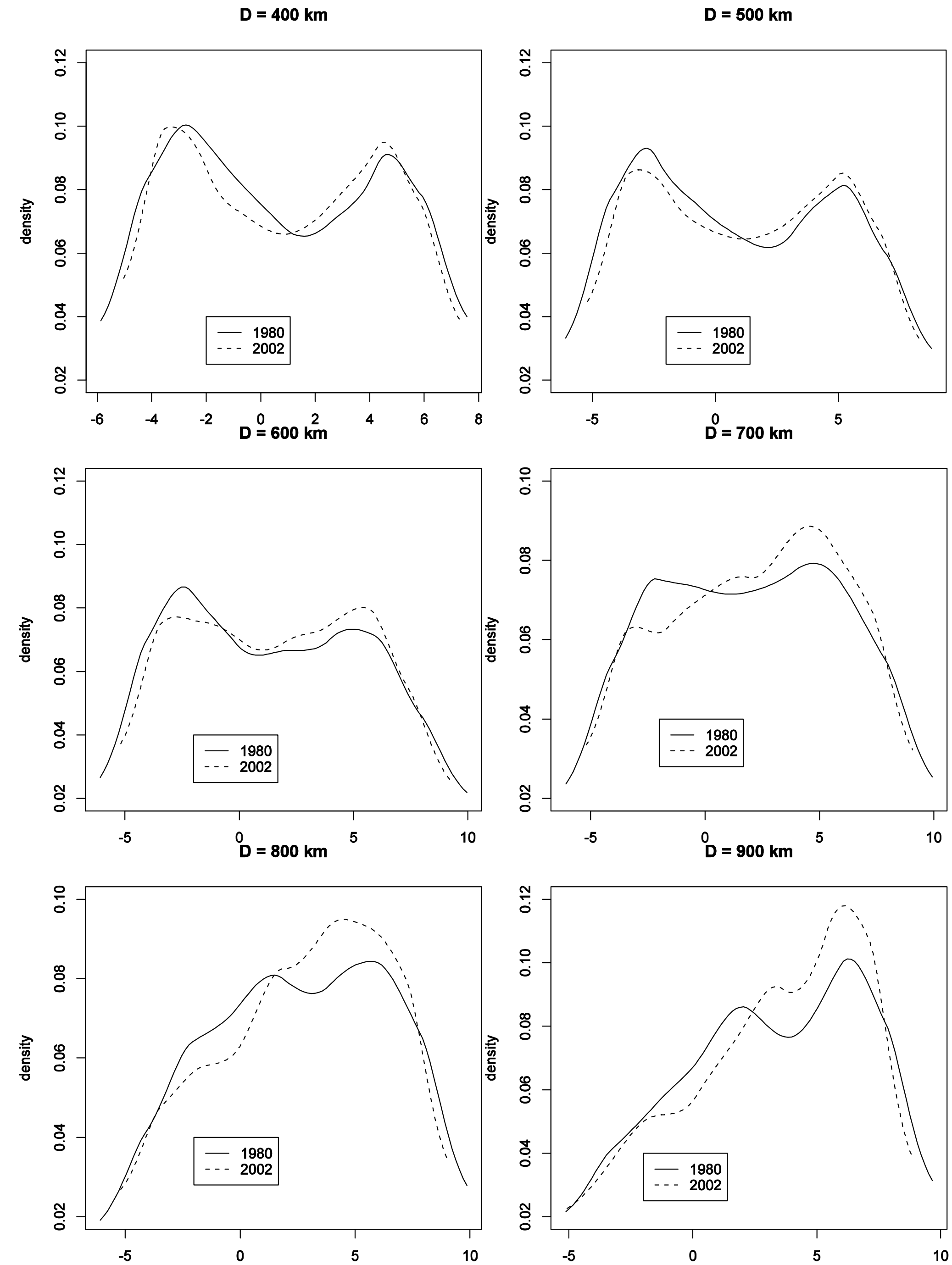
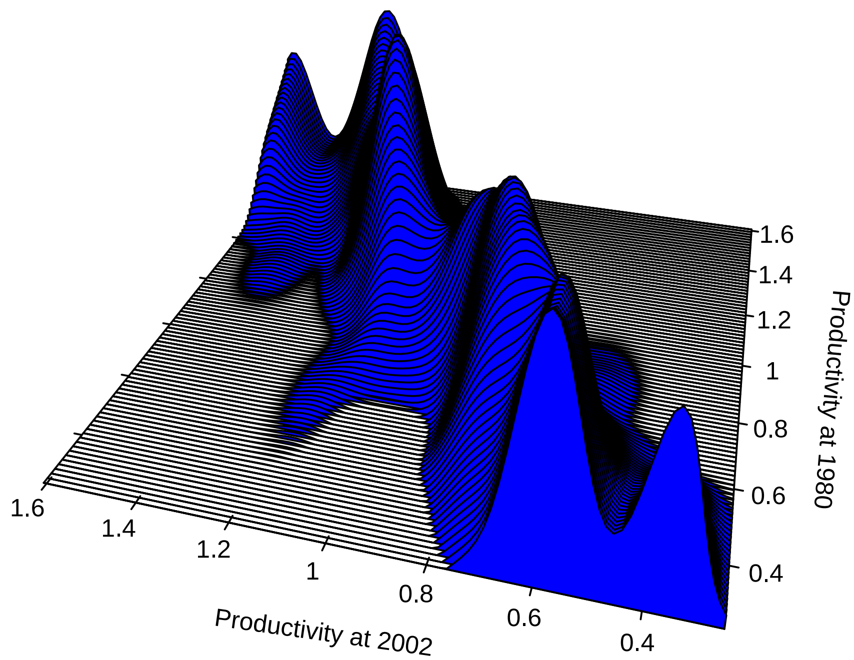
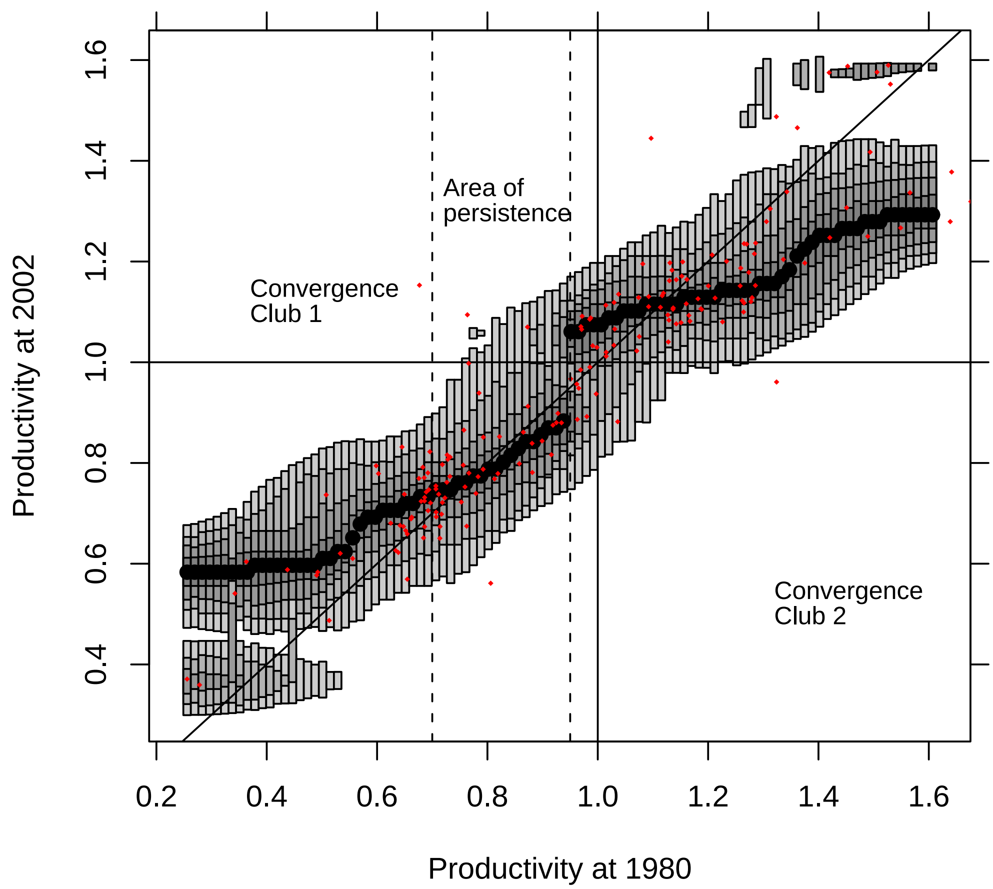


Figure 6 - Intra-Distribution Dynamics

Stacked density plot and HDR plot of conditional density based on the local parametric estimator with variable bandwidth (Hyndman and Yao, 2002)



A) Stacked density plot

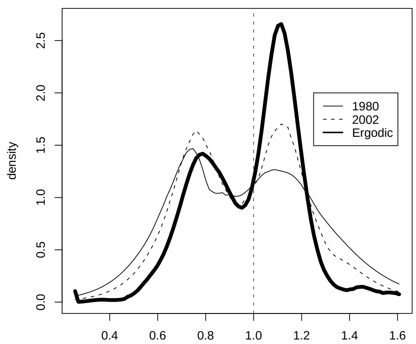


1.6
1.4
1.2
1.0
0.8
0.6
0.4
Productivity at 2002
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Productivity at 1980
Area of
persistence
Convergence
Club 1
Convergence
Club 2

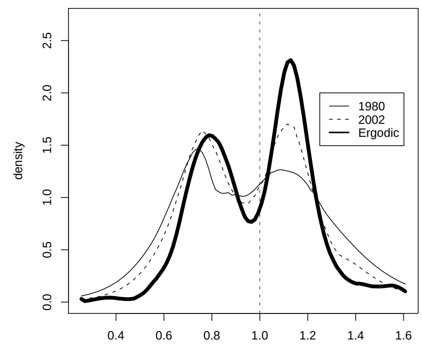
B) HDR plot

Figure 7 – Ergodic distribution conditioning on growth determinants

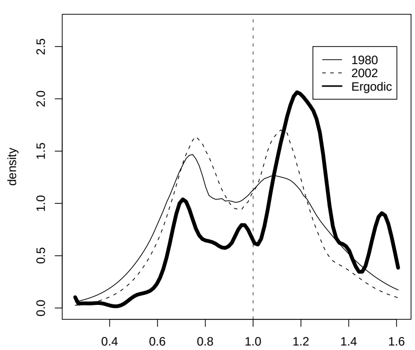
A) Unconditioned



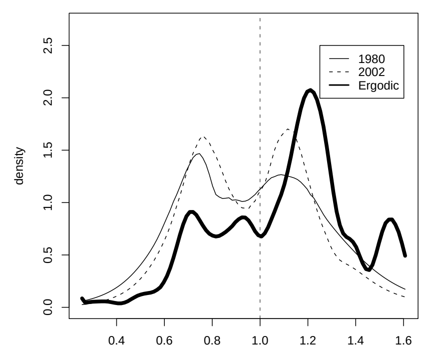
B) Conditioned on model 2



C) Conditioned on model 3



D) Conditioned on model 4



E) Conditioned on model 5

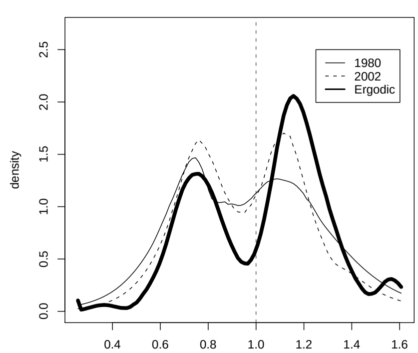
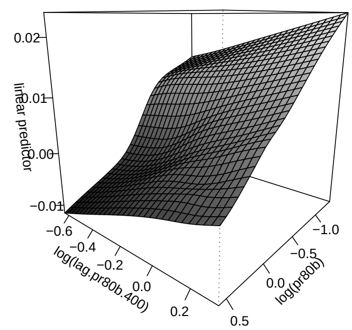
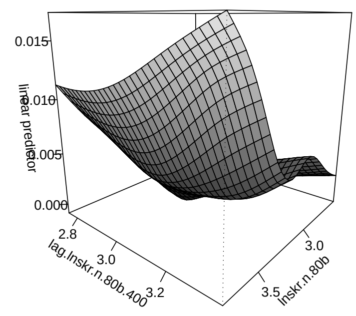


Figure 8 – Growth determinants (model 3; eq. 9)
(Distance matrix: 400 km)





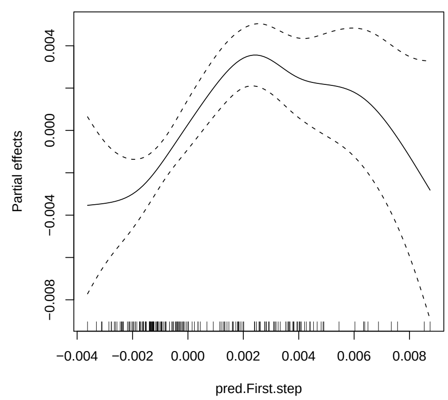


Figure 8 – Growth determinants (model 5; eq. 11)