ESTADÍSTICA (2003), 55, 164,165, pp. 261-281

REGIONAL EMPIRICS FOR ECONOMIC DISPARITIES IN ITALY: 1951-2001

GIOVANNI MARIA GIORGI
Dipartimento di Statistica, Probabilità e Statistica Applicata, University of Rome "La Sapienza"
MARIA GRAZIA PITTAU, ROBERTO ZELLI
Dipartimento di Contabilità Nazionale e APS, University of Rome "La Sapienza"

Abstract

This paper presents empirical evidence on the magnitude and evolution of economic disparities across Italian provinces over the period 1951-2001. The paper focuses on the evolution of the whole cross-sectional distributions in two perspectives. Firstly, kernel density estimates of per capita income for each decade are yielded. They reveal a persistent multimodal distribution during the whole period. The heterogeneity is subsequently modelled by a finite mixture density, whose components represent a group of poor and a group of rich provinces. Finally, the intra-distribution mobility is modelled by the stochastic kernel, following Quah (1996). The long-run distribution polarizes into two distinct classes of income. Provinces which are well off relative to national average tend to cluster around an aggregation pole, characterised by a mean value 1.2 times the national average. Provinces which are worse off tend to cluster around a mean value 30% less than the national average.

Key words: *Regional convergence; kernel estimation; mixture models, stochastic kernel*

1 Introduction

The recent years have witnessed a renewed interest of the literature in economic convergence, not only across countries but also across sub-national territorial units, giving rise to interesting debates on competing theories of spatial economic growth, as well as the adequate empirical strategies that can be used to measure the phenomenon and to compare these theories. These studies revealed that economic disparities across regions are often broader than the disparities observed across countries. This is particularly evident in Europe, where several studies documented that income differences across EU member States have fallen, but inequalities between regions within these countries have risen (see for instance Tondl, 1999; Boldrin and Canova, 2001).

The theoretical and empirical analysis of wealth disparities across Italian regions has a long tradition (see among others Tagliacarne, 1953; Lutz, 1961; Sylos Labini, 1985 and Wolleb and Wolleb, 1990) since Italy's economic development has always been characterized by territorial heterogeneity. However, due to the lack of long time-series data at a very disaggregated level, most studies focused on the economic gap between two broader territorial areas (usually the Centre-North and the South). More recently territorial disparities have been investigated at the level of administrative regions (Regioni), that corresponds to the second level of disaggregation in the Nomenclature of the Territorial Units for Statistics (NUTS) established by Eurostat, but very few attempts have been made so far to investigate territorial disparities using more disaggregated territorial units. (Notable exceptions are Ferri and Mattesini, 1996, and Fabiani and Pellegrini, 1997).

In this paper, on the basis of an original data-set recently released, we will investigate on per capita income disparities across Italian provinces, which correspond to the NUTS 3 level of classification, over the last fifty years. This territorial level of analysis seems particularly interesting for the Italian economy since the peculiar historical development of the country, where areas with brilliant performances in terms of growth are contiguous to areas with poor performances. As pointed out by Fabiani and Pellegrini (1997), the boundaries of administrative regions are the result of political and historical events which may not be related to the socioeconomic factors that form the basis of a "functional" economic area. Therefore, the use of large regions runs the risk of hiding factual spatial differences.

There is a large debate about which statistical method should be used to assess the issue of economic disparities. The traditional approach to convergence analysis is based on a cross-section regression of per capita income growth rates on initial levels of income (eventually on other determinants, as human capital, share of investments, infrastructures, ...). A negative coefficient implies that regions starting poor grow faster than regions starting rich and this is taken as $\beta-$convergence (eventually conditional on the other determinants). This measure of convergence has been criticised since the averaging out cross-section information by looking at coefficients in a regression does not provide adequate information on how regions perform relative to each other (Quah, 1996). Other downfalls of conditional β-convergence have been pointed out, among others, by Durlauf and Johnson (1995) and Lee, Pesaran and Smith (1997).

Alternatively, convergence can be analysed through the evolution of the dispersion of per capita incomes, typically measured by the standard deviation: a decreasing dispersion indicates $\sigma-$convergence. However, also this approach is not able to assess some important characteristics of the distribution, since, for example, a constant dispersion could hide mobility patterns within the distribution and persistent inequality (Quah, 1996), clearly phenomena of non convergence.

Therefore, a number of recent studies have centered on the evolution of the entire cross-sectional distributions in addressing the topic of regional convergence. The interest on distributional dynamics focuses on several aspects of regional evolutions. The first concerns the overall shape characteristics of the income distribution across provinces, and their evolution over time. A natural approach to assess these shape dynamics is to estimate the cross-section distributions by using kernel density estimation. The analysis based on kernel density, however, heavily relies on the visual impression of income distribution. To obtain a better insight into the estimated cross-sectional distributions, per-capita income can be modelled by a finite mixture density, which provides a semiparametric framework to model unknown distributional shapes through an appropriate choice of components.

The second dimension relates to the amount of internal mixing or mobility that occurs within these distributions over time. In fact, shape dynamics do not address directly the mobility pattern within the income distribution, i.e. which units move up and down the income ladder. One of the common statistical tool employed for the intra-distribution mobility analysis is the quantification of the probabilities of transition between quantiles, which gives the probability of moving between different discrete states.

Since the variables under analysis are defined on a *continuum* of values, a continuous state space process seems more appropriate to describe their behavior (see Quah, 1996; 1997). Therefore the transition function or stochastic kernel, the continuous equivalent of the transition probability matrix, has been introduced as the more adequate tool that allows to trace the evolution of the distribution of the variable, highlighting the changes in the intra-distribution mobility.

In this paper both the aspects to analyse the regional disparities across the Italian provinces will be followed. The "stylised fact" that Italy is characterised by one of the widest geographical dualism among European countries will be efficiently estimated by the kernel density and the mixture model. The behaviour of the poor and rich provinces and their relative position within the distribution will also be investigated by the stochastic kernel.

It should be noted that income disparities can be analysed considering different variables. Two measures are commonly adopted in the literature: labour productivity (GDP per units of labour) and per capita income (GDP/population). The former captures the degree of efficiency of the production system, while the latter concerns the flow of wealth to the average inhabitant and its evolution depends not only on labour productivity, but also on other factors (economic, social, cultural, demographic) that affect unemployment and participation rates (Paci, 1997). We focused the analysis on comparing economic standards of living across economies, therefore the best proxy available is the GDP per capita. Throughout the text, the terms per capita GDP or per capita income will be used as synonyms.

The remainder of the paper is organized as follows. Section 2 presents the statistical methodology of finite mixture models. In particular this Section focuses on the practical problems of estimation and selection of the minimum number of components in the mixture, compatible with the data. The empirical results of the nonparametric regional density estimates of the per capita income over time and the empirical fitting of the mixture models follow, along with a discussion on the evolution of the probabilistic clusters of poor and rich provinces generated by the components of the mixtures. In Section 3 the statistical methodology for modelling the cross-region dynamics is briefly described, together with the empirical results of the income distribution dynamics. Some conclusions are given in Section 4.

2 Modelling income distribution over time

2.1 The statistical tools: kernel smoothing and finite mixture models

From the data $x_1, x_2, ..., x_n$, a consistent estimator for the density function $f(x)$ at point x is the *fixed bandwidth* kernel estimator:

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

where $h > 0$ is the bandwidth and $K(\cdot)$ is the kernel function. The bandwidth h governs the degree of smoothness of the density estimate and the choice of the optimal bandwidth is an important empirical issue not easy to solve (Silverman 1986). Ideally, h should vary with the amount of information available, according to the data sparseness in the distribution. The *adaptive bandwidth*, that is a local smoothing, smooths out the "wiggles" in areas where the density is sparse without leading to oversmooth the dense parts of the distribution, producing an improved global estimate. The assessment of the adaptive bandwidth requires a two-step procedure. In the first step a fixed bandwidth following a statistical "optimal" rule on which a pilot estimate of the density is obtained. The second step starts with the calculation of the bandwidth weighting factors λ_i, defined as:

$$\lambda_i = \sqrt{\frac{exp\left(\frac{1}{n}\sum_{j=1}^{n} log\tilde{f}(x_j)\right)}{\tilde{f}(x_j)}} \tag{2}$$

where $\tilde{f}(x_j)$ is the pilot estimate of the density. Then, the adaptive kernel density estimate for the point x takes the form:

$$\widehat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h\lambda_i} K\left(\frac{x - X_i}{h\lambda i}\right). \tag{3}$$

The estimator (3) is the kernel estimator used in the analysis. The optimal pilot bandwidth has been chosen according to the Sheather and Jones (SJ) plug-in method (1991). The kernel function used in this analysis is the Gaussian kernel, which has the form :

$$K(\psi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\psi^2\right). \tag{4}$$

Results are not sensitive to alternative choices of the kernel function, as the Epanechnikov function.

The use of kernel densities to trace out the evolution of the income distribution provides a powerful visualization framework from which location, (multi)modality and spread can be observed simultaneously. Since the analysis based on kernel density relies to a great extent on the visual impression, in addition to this graphical approach statistical tests are also helpful to reveal certain features of the shapes that otherwise may be unmarked. For instance, Bianchi (1997) reports a bootstrap test to assess the number of modes in the density distribution of income across 119 countries. The presence of m modes can be suggestive of m unimodal underlying income distributions, each of which may refer to certain economically important region groups. In order to identify these groups, the author employees a discriminant analysis, allocating each country to one of the m groups if its per capita income falls into the interval $[c_j, c_{j+1}]$, where c_j are the cut-off points defined as values of x at which the estimated density has a local minimum.

An alternative way to characterize these groups can be achieved by means of a mixture of distributions with different mixing proportions. A mixture model of finite distributions represents an extremely flexible method in modelling heterogeneous distributional shapes, apparently unable to be modelled by a single component distribution. In fact, multimodality in the income distribution does not necessary result in the same number of distinct underlying univariate components, especially when the component are not sufficiently far apart. On the other hand, when the distribution is unimodal, a finite mixture might still be appropriate when, for example, the components have similar means and large variances. Another advantage in implementing this approach is that the fitting of the finite mixture model provides a probabilistic clustering of the sample units in terms of their conditional probabilities of membership of each components of the mixture.

More formally, with a mixture model-based approach, the observed data set can be viewed as having come from a mixture of g component distributions in unknown proportions $\pi_1, ..., \pi_g$, where the mixing proportions π_i are nonnegative and sum to one. In detail, the probability density function of the random vector X_j under a g-component mixture model is defined in parametric form as:

$$\Phi(x_j, \Psi) = \sum_{i=1}^{g} \pi_i \Phi_i(x_j, \theta_i), \tag{5}$$

where the vector $\Psi = (\pi_1, ..., \pi_{g-1}, \xi')'$ contains all the unknown parameters in the mixture model and the vector ξ contains all the parameters $(\theta_1, ..., \theta_g)$ known *a priori* to be distinct; $\Phi_i(x_j, \theta_i)$ denotes the univariate density evaluated at x_j depending on a parameter vector θ_i; and π_i, $i = 1, ..., g$ represent the mixing proportions. While the mixing parameter π_i gives the prior probability that the region belongs to the ith component of the mixture, representing an *endogenous* parameter which determines the relative importance of each component in the mixture distribution, the posterior probability τ_{ij} given by:

$$\tau_{ij} = \tau_i(x_j, \Psi) = \frac{\pi_i \Phi_i(x_j, \theta_i)}{\sum_{i=1}^{g} \pi_i \Phi_i(x_j, \theta_i)}, \tag{6}$$

represents the probability that the jth province with income x_j comes from the ith component of the mixture.

The fitting of (5) provides a partition of the data directly derived from the Maximum Likelihood (ML) estimates of the mixture parameters by assigning each province with income x_j to the component which provides the greatest conditional probability (6) that x_j arises from it. In the mixture approach, the parameter Ψ is chosen to maximize the log-likelihood function in (5). However, due to the the stressed out computational difficulties, (see, *inter alia* Aitkin and Wilson, 1980), the log-likelihood in the mixture models is difficult to maximize directly. The EM algorithm (Dempster *et al.*, 1977) provides a simple iterative method for the ML estimates in a wide class of "missing data" problems, including general mixture distributions. Starting form an initial parameter $\Psi^{(0)}$, an iteration of the EM algorithm consists in computing the current conditional probabilities $\tau_i(x_j, \Psi)$, $(j = 1, 2, ..., n; i = 1, 2, ..., g)$ that x_j arises from the i-th component of the mixture for the current value of Ψ, according to the equation (6) (E-step). Then, the ML estimates are computed using the conditional probabilities $\tau_i(x_j, \Psi)$ as conditional mixing weights (M-step). The sequence of alternate E and M steps continues until convergence occurs to the ML estimates. The algorithm is easy to implement and it converges relatively fast. It yields the ML estimates of the parameters and the maximized log-likelihood function from which a primary idea of the number of components in the mixture can be achieved. When there is no *a priori* information, a more formal way to identify the distributions underlying the mixture model is to test for the number of components membership in the model. After the maximized log-likelihood function has been evaluated, it is possible to formulate likelihood ratio tests for testing the null hypothesis $H_0 : g = g_0$ against the alternative $H_1 : g = g_1$, for some $g_1 > g_0$. As well known in the literature, the asymptotic null distribution of the Likelihood Ratio Test Statistic (LRTS), $-2\log\lambda = 2\left\{\log L(\widehat{\Psi}_1) - \log L(\widehat{\Psi}_2)\right\}$ is not necessarily be a chi-squared, being $L\left(\widehat{\Psi}_0\right)$ and $L\left(\widehat{\Psi}_1\right)$ the likelihood of the mixture model, respectively, under the null and under the alternative hypothesis. The choice of the number of components, g, can be assessed by a bootstrap test (McLachlan, 1987). The resampling approach is based on the null distribution of $-2\log\lambda$ for $H_0 : g = g_0$ versus $H_1 : g = g_1$. The LRTS is bootstrapped as follows. Under the null hypothesis of g_0 components in the original sample, B bootstrap samples are generated parametrically. For each bootstrap sample, the mixture model is fitted for both $g = g_0$ and $g = g_1$ and consequently the quantity $-2\log\lambda$ is evaluated. The null distribution of $-2\log\lambda$ can be finally estimated from B replicated values of $-2\log\lambda$. Let $\widehat{\theta}^*(b) = -2\log\lambda(b)$, $b = 1, 2, ..., B+1$ be the ordered bootstrap replications of $-2\log\lambda$, inclusive of the value of $-2\log\lambda$ evaluated in the original sample $\widehat{\theta}^*$ and suppose that $\widehat{\theta}^*$ is in the rth position of the ordered list of the replications. Therefore, the estimate of the achieved significance level P, corresponding to the value $-2\log\lambda$ evaluated from the original sample, is:

$$P - value \simeq 1 - \frac{r}{B+1}. \tag{7}$$

The class of finite mixtures that will be considered for describing the income distribution in Italy is the class of finite normal mixture. The provincial data are considered as arising from a mixture of g normal distributions in unknown proportions $\pi_1, ..., \pi_g$:

$$\Phi(x_j, \Psi) = \sum_{i=1}^{g} \pi_i N\left(\mu_i, \sigma_i^2\right) \tag{8}$$

where $\theta_i = (\mu_i, \sigma_i^2)$ denotes the mean and the variance of the i-th univariate normal component and π_i the respective mixing proportion. The vector of the all unknown parameters which will be estimate by maximum likelihood is $\Psi = \left(\pi_1, ..., \pi_{g-1}, \mu_1, ..., \mu_g, \sigma_1^2, ..., \sigma_g^2\right)'$.

The assumption of normality could be restrictive, since theoretically any functional form can be considered. However, the family of normal mixture densities of g component is very flexible and a wide variety of density shapes can be approximated by a normal mixture (Marron and Wand, 1992). Typical of mixture models is the presence of multiple local maxima in the likelihood function (Aitkin and Wilson, 1980; McLachan and Peel, 2000). Particularly, in mixtures of univariate normal components with unequal variances, the likelihood function is unbounded and which root of the likelihood equation to choose as the estimate is not straightforward. The associated practical problem regards the selection of different starting values for the EM algorithm. In this paper, following by McLachan *et al.* (1999), randomly selected starts, hierarchical clustering-based starts and k-means clustering-based starting values have been selected as different starting values for the vector Ψ of unknown parameters and for the posterior probabilities.

2.2 Time evolution of regional disparities: are twin peaks persistent?

The series of provincial GDP has been recently released by the National Statistical Office (Istat) after the implementation of the new System of National Accounts (ESA 95) and only for the years 1995-1999. However, the Istituto Tagliacarne, the research centre of the Union of the Italian Chambers of Commerce, has recently provided an homogeneous series of per capita income at provincial level since 1951 (Unioncamere, 2001).

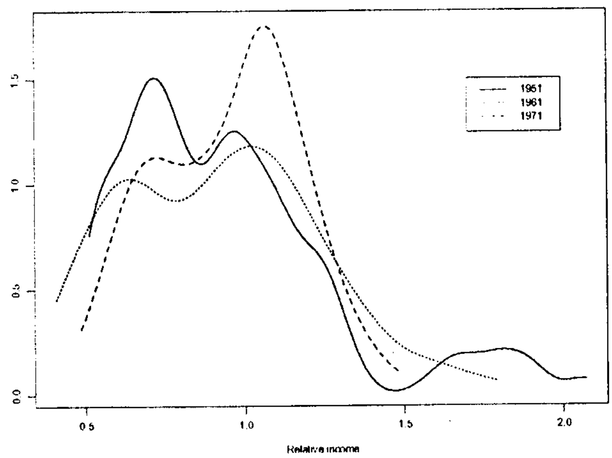
In taking per-capita GDP as a measure of development, there are some drawbacks to be aware of. It should be noted that GDP is a workplace-based measure of economic activity in an area and it is not equivalent to the final disposable income of private households resident in a given province. Therefore it can be considered only a rough approximation of the economic capacity of individuals to acquire goods and services. Unfortunately, a time series of per capita disposable income at NUTS 3 level is not available yet.

The income level in each province has been divided by that for Italy, so that the analysis is performed using the per capita income in the provinces relative to the (population weighted) average. This transformation of the data removes the common growth and the business cycle effect and circumvents the problems related to the deflation of the current values.

Since evolution is a slow process, large movements in the distributions are not expected year-on-year. Therefore, to better catch greater changes in the static distributions, income per-capita distribution for $n = 95$ Italian provinces is measured in 1951, 1961, 1971, 1981, 1991 and 2001. For these years the estimation of income is also more reliable since it has been estimated consistently with the population and the industrial censuses. The GDP has been estimated according to the previous System of National Accounts (ESA 79). It should be noted that over the period of the analysis, the number of provinces has been changed. The series available refers to the number of provinces before 1992, when 8 new provinces have been created.

The evolution of the static distributions over time and significant changes in the shape of the distributions over this 50-year period can be caught by the inspection of the densities. Figures 1 and 2 report the adaptive kernel estimates of the relative income densities for the selected years. Table 1 reports the results of the bootstrap test for the choice of the number of components in the mixture model. The results show that, apart from the year 1951, the remainder years does not reject the hypothesis of $g = 2$ components. Table 2 shows for each year the estimates of the means, variances and mixing proportions of the previously selected

Figure 1: Adaptive kernel estimate of per capita income in the Italian provinces, years 1951, 1961, 1971

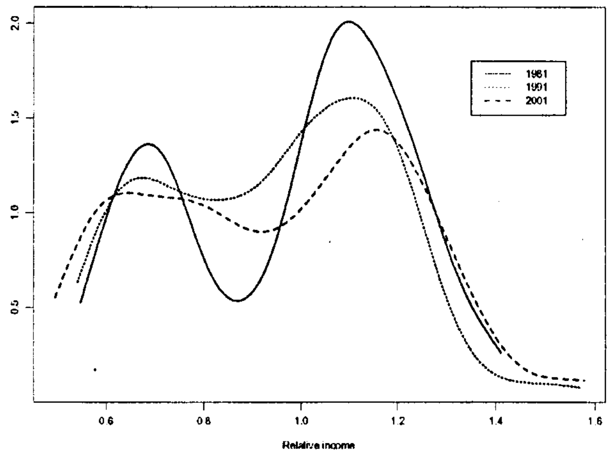


components of the fitted mixture model.

The estimated density for 1951 shows two well separated peaks, the highest corresponding at a value below the average and the other one just above it. Moreover, it seems to appear another possible peak in the right tail of the density corresponding to the richest regions. The bootstrap test for the number of normal components of the mixture confirms $g = 3$ components. Figure 3 shows the kernel density estimation, the fitted model and the single components for 1951. Looking at the posterior probabilities $\widehat{\tau}_{ij}$, in the wealthiest cluster are included, with a probability higher than 0.9, provinces belonging to the most industrialized part of the country (the so-called Milano-Torino-Genova triangle). Several backward provinces of the North-East (like Udine, Rovigo, Belluno Pordenone, Padova) and the almost all the Southern provinces belong, with high probability, to the "poorest" cluster, with some notable exceptions (Matera, Foggia, Siracusa). The mixing proportion associated to this cluster is equal to 39.7%, indicating a consistent percentage of provinces belonging to this component.

One decade later, the "lump" in the upper tail of the distribution found in 1951 seems to vanish and also the bimodality presents different features. In fact, the two peaks have almost the same height and seem more separated, accentuating in this way the imbalance between the two groups of provinces generated by the two components of the mixture, as confirmed by the bootstrap test (see Table 1), even though the mixing proportion of the first component (26.1%) indicates a reduced percentage of "poor" provinces. In terms of conditional probabilities, it should be noted that, apart from Frosinone and Rieti (two provinces in the Centre), only Southern provinces belong, with a high probability, to this group. All the Center-Northen areas, conversely, are included in the more prosperous group. This evidence

Figure 2: Adaptive kernel estimate of per capita income in the Italian provinces, years 1981, 1991, 2001



suggests that in the early sixties the dualism between the Centre-North and the South is more pronounced.

The results for the year 1971 confirm the bimodality in the distribution, even though it is less evident, and the presence of two normal components. The percentage of the "poor" provinces, given by the mixing proportion of the first component, is almost unchanged and there are no significant movements of the individual provinces between the components of the distribution. Although the difference between the means of the relative per capita income of the two groups has decreased, the tendency of converging does not seem very remarkable.

In fact, in 1981, the two modes of the distribution appear again well separated and the gap between the two components seems wider, as confirmed by the means of the poor and the rich group (see also Figure 4). These two groups can be well identified, since the conditional probability of each province clearly allows to allocate it in one of the two groups. The composition of the two groups reveals a persistent dualism between the two broader areas of the country.

In the year 1991, the twin-peaks are still present and a two-component mixture is compatible with the data. The position of the first mode does not change, but its shape modifies noticeably in that it gets smaller and wider. Consequently, it is more difficult, with respect to 1981, to detect a well-defined group of poor provinces and a group of wealthier provinces. The same comment can be extended for the last year of the analysis. The first mode is small and wide, and, differently from 1991, the second mode shifts to the right by a significant amount, showing an increasing separation of the two modes and of the two underlying components of the mixture (see Figure 5).

Overall, per capita income tend to spread according to a two-peaks distribution and the

Table 1: The choice of the number of components.

	AIC	BIC	-2logλ	*P-value*		AIC	BIC	-2logλ	*P-value*
g	1951				g	1981			
1	67.32	72.42	-	-	1	-3.43	1.68	-	-
2	41.43	54.20	31.88	0.00	2	-39.51	-26.75	42.09	0.00
3	35.46	55.89	11.97	0.06	3	-36.31	-15.87	2.79	0.81
g	1961				g	1991			
1	48.00	53.11	-	-	1	-4.90	0.21	-	-
2	41.87	54.64	12.13	0.03	2	-16.45	-3.68	17.55	0.01
3	43.99	64.42	3.88	0.59	3	-14.95	5.48	4.50	0.49
g	1971				g	2001			
1	-10.06	-4.95	-	-	1	22.20	27.31	-	-
2	-15.19	-2.42	11.13	0.05	2	9.75	22.52	18.45	0.00
3	-17.94	2.49	8.76	0.14	3	6.04	26.47	9.70	0.11

occurrence of a two-component mixture model suggest the presence of a polarisation process in the per capita income distribution. Almost all the Southern provinces are included in the poorest cluster of provinces for each year of the analysis.

It should be noted that the fitted mixture model can be instead interpreted in a dynamic context, in fact the switching from one group to another one depends on the relative movement of a province in the distribution with respect to other provinces. Therefore to evaluate the intra-distribution mobility of new wealth, the immediate strategy is to examine the switches of provinces from the "poor" and the "rich" group and vice versa.

From Table 2, we observe that the mixing proportion associated to the "poorest" component has a fluctuating behavior and it does not show a tendency to reduce over time.

To analyse the mobility between groups, it is necessary to allocate the individual provinces

Table 2: Means, variances and mixing proportions of the fitted mixture model.

	Mean			Variance			Mixing proportion		
years	g_1	g_2	g_3	g_1	g_2	g_3	g_1	g_2	g_3
1951	0.665	1.019	1.797	0.010	0.026	0.023	0.397	0.519	0.084
1961	0.579	1.044	-	0.009	0.067	-	0.260	0.740	-
1971	0.675	1.059	-	0.009	0.025	-	0.291	0.709	-
1981	0.682	1.120	-	0.005	0.015	-	0.323	0.677	-
1991	0.645	1.040	-	0.004	0.032	-	0.236	0.764	-
2001	0.672	1.124	-	0.011	0.030	-	0.376	0.624	-

Figure 3: Distribution of relative income in 1951: kernel density estimate and ML estimate of mixture of three normal components

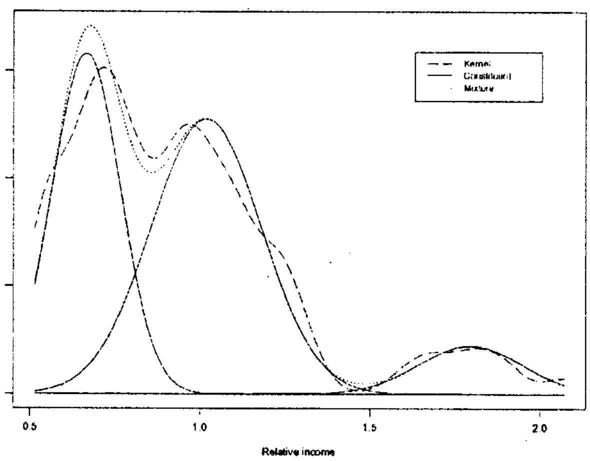


Figure 4: Distribution of relative income in 1981: kernel density estimate and ML estimate of mixture of two normal components

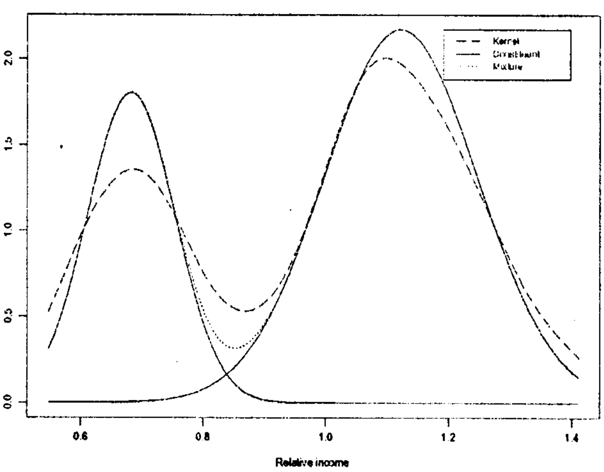
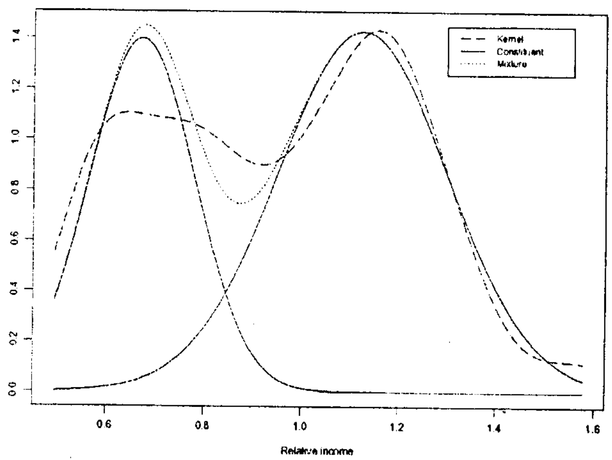


Figure 5: Distribution of relative income in 2001: kernel density estimate and ML estimate of mixture of two normal components



in one of the groups. The *ex post* estimated probabilities are helpful for this identification. In this analysis we chose to label a province j as "poor" if its conditional probability with respect to the first component of the mixture, $\hat{\tau}_{1j}$, is larger than 0.70. If $0.70 < \hat{\tau}_{1j} < 0.30$ we decided that the province can not be clearly allocated in one of the two groups, while for $\hat{\tau}_{1j} \leq 0.30$ we labelled the province j as "rich". Table 3 reports, for each year, the number of provinces that belong to the "poor" group, to the "rich" group and those provinces not clearly allocated in one of these two groups. For a more homogeneous comparison, the third group of the richest provinces, that is present only in the first year, is non considered in this table.

As reported in Table 3, in the period 1951-2001 the number of "poor" provinces drops from 36 to 34, while the number of the "rich" provinces increases from 52 to 56. The not unequivocally allocated provinces change from 7 to 5. Furthermore, a fall of the number of the "poor" provinces and a consecutively increase of the number of rich provinces can be observed in 1991. The "poor" provinces are only 20, the "rich" are 69 representing the highest number of the "rich" provinces in the fifty years period.

To better investigate the intra-distribution movements of the Italian provinces, it may be useful to extend the number of groups and to examine the mobility within these groups. In this view, a suitable approach is the analysis of the probabilities of transition between quantiles or in general discrete states, as Quah (1993) suggested. An obvious problem is the choice of the discretization since the variable under analysis is defined on a continuum of values. Therefore a continuous state space process seems more adequate to analyse its behavior (see Quah, 1996; 1997). The continuous counterpart of the transition probability matrix is the transition function or stochastic kernel,which has been adopted in the recent

Table 3: Number of "poor" and "rich" provinces based on their *ex post* probabilities.

years	1951	1961	1971	1981	1991	2001
"poor"	36	23	26	31	20	34
"rich"	52	63	60	64	69	56
not allocated	7	9	5	0	6	5
total provinces	95	95	95	95	95	95

literature (see Johnson, 2000).

In the next section, the stochastic kernel will be introduced in a more formal way, while the main results of the intra-distribution dynamics of the Italian provinces will be presented subsequently.

3 The intra-distribution mobility

3.1 Measuring mobility (via the stochastic kernel)

The underlying theoretical framework of the stochastic kernel can be summarised as follow (a far more general formalization can be found in Quah (1997)). Let X_t the relative income of the Italian provinces at time t and ϕ_t its probability distribution. The simplest scheme to describe the evolution of the distribution of incomes across provinces is analogous to the first order autoregression from standard time-series analysis:

$$\phi_{t+1} = T^* (\phi_t, u_{t+1}) = T^*_{u_t} (\phi_t) \,, \mathrm{t} \geq 1, \mathrm{given}\phi_0 \tag{9}$$

where the operator T^* maps the distribution from period t to period $t+1$, u_{t+1} is a sequence of disturbances and $T^*_{u_t}$ absorbs the disturbance into the definition of the operator. We focus on the estimation of the operator T^* which describes the evolution of the distribution of incomes across regions.

If X_t is a discrete income-space variable, the operator T^* can be interpreted as the transition probability matrix M_t of a Markov process and (9) becomes:

$$\phi_{t+1} = M_t^{'} (\phi_t) \,. \tag{10}$$

If the further assumption of time-invariance of the underlying transition mechanism is introduced, the matrices M_t's can be averaged obtaining a single transition probability matrix M describing the dynamics of the discretized distribution.

Under the two assumptions of first order evolution and time invariance the equation (9), $\tau - 1$ periods later, becomes:

$$\phi_{t+\tau} = (M^\tau)^{'} (\phi_t) \,. \tag{11}$$

Since M is a transition probability matrix, its largest eigenvalue is 1, and the left eigenvector corresponding to that eigenvalue has all entries nonnegative and summing to 1. Since in general the largest eigenvalue is unique, M^{τ} converges to a rank-one transition matrix. But then all its rows must be equal, and moreover equal to that probability vector satisfying:

$$\phi_{\infty} = M' \phi_{\infty}. \tag{12}$$

The vector ϕ_{∞} is the *ergodic* row vector, corresponding to the limit of (10) as $t \longrightarrow \infty$ and representing the long-run limit of the distribution of incomes across provinces.

However, more often the income variable X_t assume infinite values becoming a continuous income-space variable. In this case the operator T^* must be interpreted as a transition function or stochastic kernel. Furthermore, as pointed out by Quah (1997), the arbitrariness of the discretization in the income state-space can introduce a bias in the dynamic analysis altering the conclusions. The stochastic kernel can be therefore viewed as the appropriate generalization of a transition probability matrix: $f(x, t, y, \tau)$ giving the p.d.f. of $X_{t+\tau} = y$ conditional on $X_t = x$. Thus:

$$prob\,(a < X_{t+\tau} < b | X_t = x) = \int_a^b f(x, t, y, \tau)\, dy. \tag{13}$$

If the transition mechanism is time invariant, the transition probability density depends only on the time interval τ and can be written $f_\tau(y|x)$, giving the p.d.f. of $X_{t+\tau} = y$ conditional on $X_t = x$, $\forall t$.

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

$$f_{t+\tau}(y) = \int_0^{\infty} f_\tau(y|x) f_t(x) dx, \tag{14}$$

where $f_\tau(y|x)$ is the density of y conditional on an income x, τ periods later.

As in the discrete income state space, the long-run limit of the distribution of incomes across regions is the limit of (14) as $t \longrightarrow \infty$. The resulting ergodic distribution is (Johnson, 2000):

$$f_{\infty}(y) = \int_0^{\infty} f_\tau(y|x) f_{\infty}(x) dx. \tag{15}$$

The estimation of (15) requires the estimation of the transition matrix $f_\tau(y|x)$ that can be obtained by dividing the estimated joint density by the estimated marginal density.

The stochastic kernel has been estimated for a 10-year transition period, setting $\tau =$ 10. The relative income data, estimated for 1951, 1961, 1971, 1981, 1991 and 2001 give observations on five 10-year transitions for each province to estimate $f_{10}(y|x)$.

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

$$\widehat{f}_{t,t+\tau}(x,y) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_x}, \frac{y-y_i}{h_y}\right)}{nh_x h_y}, \tag{16}$$

where $z_i = (x_i, y_i)$ is the ith sample observation and $z = (x, y)$ is a fixed point.

The kernel function used in the estimation of (16) is the standard bivariate Gaussian kernel. The standard normal kernel is a "product kernel" which is the product of two univariate kernel and this is the kernel that minimizes the mean integrated squared error ($MISE$) of the estimator over the class of product kernel (Pagan and Ullah, 1999).The optimal bandwidths h_x h_y·have been chosen according to the Sheather and Jones plug-in method (1991).

Given the joint distribution, the marginal distribution of x has been derived by numerically integrating the joint distribution with respect to y.

$$\widehat{f}_t(x) = \int_y \widehat{f}_{t,t+\tau}(x,y)dy = \int_y K(x,y)dy. \tag{17}$$

Finally, the conditional distribution has been obtained dividing the joint distribution by the marginal one:

$$\widehat{f}_\tau(y|x) = \frac{\widehat{f}_{t,t+\tau}(x,y)}{\widehat{f}_t(x)}. \tag{18}$$

3.2 Distribution dynamics across Italian provinces

Figure 6 reports the three dimensional stochastic kernel and Figure 7 the relative two dimensional contour plot. Cutting horizontally the stochastic kernel, the conditional distribution of the relative income at time $t + 10$ given its value at time t can be obtained. In fact, stand at any point on the period t axis, and then look in a straight line, parallel to the period $t+10$ axis, the surface of the kernel is traced out, obtaining a probability density that is always non-negative and integrates to 1. The more likely transition probabilities manifest as higher values on this line. The 45-degree diagonal indicates persistence properties and when most of the graph is concentrated along this line, then the elements in the distribution remain where they started. As evident from Figure 6, a large portion of the probability mass remains clustered along the diagonal, but some peaks lie above and below the 45° diagonal, respectively for low income and for high income ranges. This characteristic implies that the provinces with incomes below the average tend to increase their relative incomes and the provinces with incomes above the national average tend to decrease their relative incomes over the 10-year horizon. This feature is more clearly shown in Figure 7 which plot the contour of the surface of Figure 6.

Although the remarkable stability in the middle income group, ranging from about 0.5 to 1.5 times the Italian average income, important changes happen in the lower (less than 0.5 times the Italian average income) and in the upper (more than 1.5 the Italian average income) income group, as clearly marked by the peaks above and below the 45° line, the provinces with very low incomes tend to move up in the relative distribution, even though the magnitude of this tendency to grow is very low. The mobility pattern characterizing the rich provinces is instead more significant, as the small peaks below the 45° line in Figure 7 reveal. The peaks below the diagonal also indicate that the rich provinces tend to move down in the relative distribution, becoming poorer. Consequently, the catching up with the Italian provinces could be interpreted as a sort of "catching down". Nonetheless, the convergence process within Italian provinces is very slow and the "two peaks" distribution persists. This feature appears stronger in the ergodic distribution, reported in Figure 9, in which the two groups of provinces are unequivocally displayed. The ergodic density function, obtained solving equation (15), describes in fact the income distribution across Italian provinces achieved in the long-run under the hypotheses first order evolution and time invariance. The estimated ergodic distribution is highly bimodal with two "long-run tendency clubs", one below and the other one above the national average. In fact, the first peak occurs at around the value of 0.7, while the second one falls around the value of 1.2 times the national average, indicating a process of polarization of the Italian provinces, where the majority of it tend to locate in the wealthier group, while a relative smaller number of provinces persistently stay poor. As suggested by Johnson (2000), additional information on the distribution dynamics can be yielded by plotting the estimated median value of relative income at time $t+10$ conditional on its value at time t. When the median line lies above the 45° diagonal line, the income median at time $t+10$ exceeds the actual value at time t, revealing a tendency for relative income provinces to rise with the passage of time. As shown in Figure 8, the conditional median at time $t+10$ exceeds its value at time t only for the very low relative income range exhibiting the tendency of poorest regions to catch up. For the high relative income values, the conditional median at time $t+10$ lies above its value at time t, confirming the tendency of the richest provinces to move down the relative distribution.

Figure 6: Relative Income Dynamics across 95 Italian Provinces: the estimated $f_{10}(y|x)$

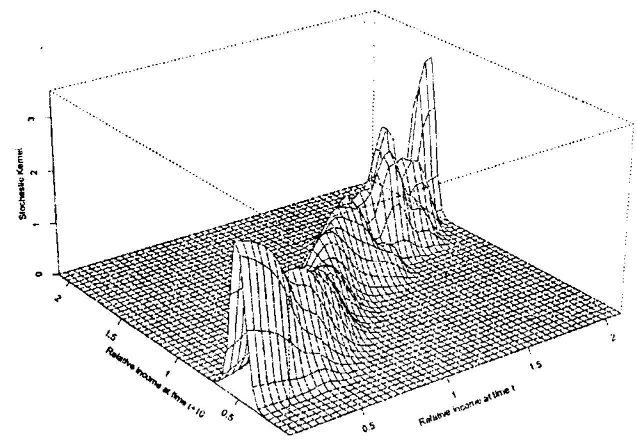


Figure 7: Relative Income Dynamics across 95 Italian Provinces: Contour Plot of estimated $f_{10}(y|x)$

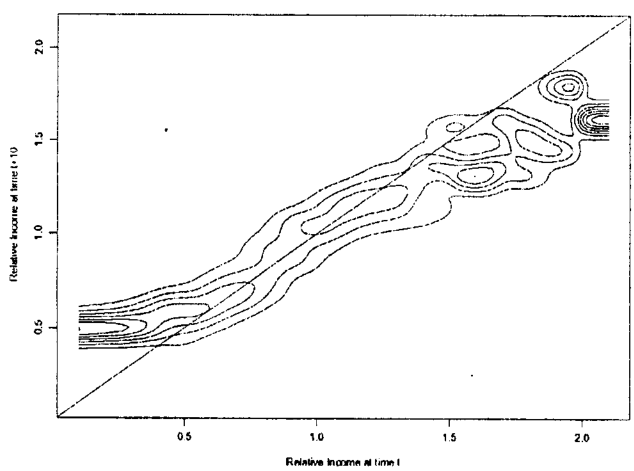


Figure 8: Median income at time $t + 10$ conditional on income at time t

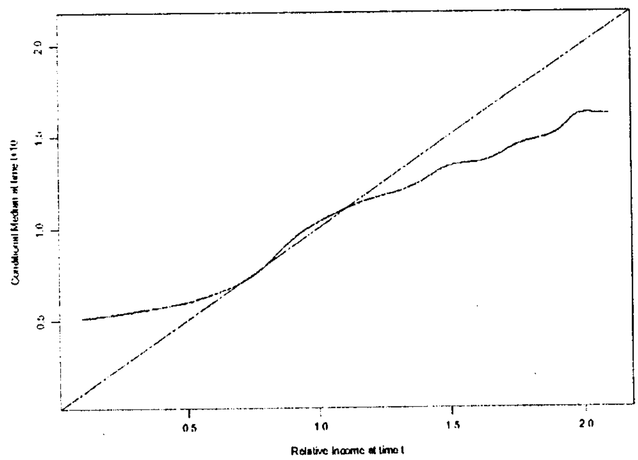
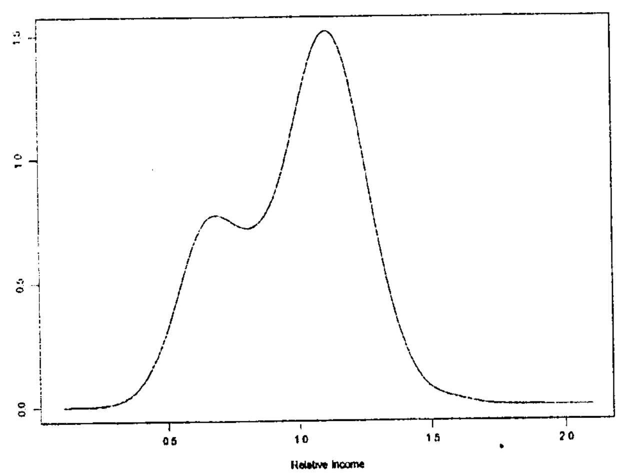


Figure 9: Estimated ergodic density



4 Concluding remarks

The examination of the NUTS 3 data of income per capita in Italy shows that there is no tendency for the distributions to have an unimodal representation. In fact, the twin-peaks feature of the distributions seems to be persistent. For example, the spread between the mean and the variance of the two selected components of the mixture distribution is largely unchanged over time.

Furthermore, the intra-distribution analysis of income per capita denies that fast convergence in levels occurs across the Italian provinces. It also suggests the idea that divergence and polarization exists. More specifically, we do not notice any form of systematic "catching-up" of poor provinces, but eventually a phenomenon of "catching down" emerges. That is, it seems that there is a tendency for territorial units which started well above the mean to regress toward the national average.

Under the hypothesis of time invariance, the long run distribution evolves towards bi-modality. Provinces which are well off relative to national average tend to cluster around an aggregation pole (characterised by a mean value 1.2 times the national average), while provinces which are worse off tend to cluster around another aggregation pole (characterised by a mean value 30% less than the national average). This tendency would be consistent to the club convergence hypothesis, reflecting a persistent heterogeneity in the Italian economic structure.

After extracting the relevant data, all computations in this paper were performed using S-Plus.

5 Acknowledgements

Financial support provided by MIUR is gratefully acknowledged. Maria Grazia Pittau would like to thank Paul Johnson for his precious support and for his helpful comments, while Roberto Zelli is grateful to Istituto Tagliacarne and in particular to Alessandro Rinaldi for suggestions and for making the provincial data set available.

References

AITKIN, M. and WILSON, G. T. (1980). "Mixture models, outliers, and the EM algorithm". *Technometrics*, **22**: 325-331.

BERNARD, A.B. and DURLAUF, S.D.(1995). "Multiple regimes and cross-country growth behaviour, convergence in international output". *Journal of Applied Econometrics*, **10**: 97-108.

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

BOLDRIN, M. and CANOVA, F. (2001). "Inequality and convergence in Europe's regions: reconsidering European regional policies". *Economic Policy*, **32**: 207-253.

CHESHIRE, P.C. and CARBONARO, G. (1995). "Convergence-divergence in regional growth rates: an empty black box?". In: Vickerman R.W., Armstrong H.W. (Eds), *Convergence and Divergence among European Regions*, Pion, London.

DEMPSTER A.P., LAIRD N.M. and RUBIN D. B. (1977). "Maximum likelihood from incomplete data via EM algorithm". *Journal of the Royal Statistical Society*, B, 69: 1-38.

DIGNAN, T. (1995). "Regional disparities and regional policy in the European Union". *Oxford Review of Economic Policy*, **11**: 64-95.

DURLAUF, S., and JOHNSON, P.A. (1995). "Multiple regimes and cross-country growth behavior". *Journal of Applied Econometrics*, **10**: 365-84.

FABIANI, S. and PELLEGRINI, G. (1997). "Education, infrastructure, geography and growth: an empirical analysis of the development of Italian provinces". *Temi di Discussione* , n.**323**, Banca d'Italia, Roma.

FERRI, G. and MATTESINI, F. (1996). "Accounting for local economic growth: some postwar evidence on Italian provinces". Istituto di Studi Economici, Quaderni ISE n. 96.

FINGLETON, B. (1999). "Estimates of time to economic convergence: an analysis of regions of the European Union". *International Regional Science Review*, **22**: 5-34.

JOHNSON, P.A. (2000). "A nonparametric analysis of income convergence across the US states". *Economics Letters*, **69**: 219-223.

LEE, K., PESARAN, M.H. and SMITH, R. (1997). "Growth and convergence in multi-country empirical stochastic Solow model". *Journal of Applied Econometrics*, **12**: 357-392.

LUTZ, V. (1961). "Alcuni aspetti strutturali del problema del Mezzogiorno: la complementarità dell'emarginazione e dell'industrializzazione". *Moneta e Credito*, **14**: 407-443.

MAGRINI, S. (1999). "The evolution of income disparities among the regions of the European Union". *Regional Science and Urban Economics*, **29**: 257-281.

MARRON, J. S. and WAND, M. P. (1992). "Exact mean integrated squared error". *The Annals of Statistics*, **20**: 712-736.

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

MCLACHLAN, G.J., PEEL, D., BASFORD, K.E., and ADAMS, P. (1999). "The EMMIX software for the fitting of mixtures of normal and t-components". *Journal of Statistical Software*, **4**: issue 2.

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

NEVEN, D. and GOUYETTE, C. (1995). "Regional convergence in the European Community". *Journal of Common Market Studies*, **33**: 47-65.

PACI, R. (1997). "More similar and less equal: economic growth in the European regions, *Weltwirtschaftliches Archiv*, **133**: 609-634.

PAGAN, A. and ULLAH, A.(1999). *Nonparametric Econometrics.* Themes in Modern Econometrics, Cambridge University Press, Cambridge.

QUAH, D. (1993). "Empirical cross-section dynamics in economic growth". *European Economic Review*, **37**: 426-434.

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

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

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

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

SYLOS LABINI, P. (1985). "L'evoluzione economica del Mezzogiorno negli ultimi trent'anni". *Temi di Discussione*, n.**46**, Banca d'Italia, Roma.

TAGLIACARNE, G. (1953). "Calcolo del reddito del settore privato e della pubblica amministrazione nelle province e regioni d'Italia nel 1952 e indici del potere d'acquisto e del risparmio bancario e postale". *Moneta e Credito*, **22**: 1-40.

TONDL, G. (1999). "The changing pattern of regional convergence in Europe". *Review of Regional Research*, **19**: 1-33.

UNIONCAMERE, (2001). "Spunti per una lettura di cento anni dello sviluppo territoriale italiano 1901-2001". Centenario dell'Unione Italiana delle Camere di Commercio, Roma 13 Dicembre 2001.

WOLLEB, E., and WOLLEB, G. (1990). *Divari regionali e dualismo economico.* Il Mulino, Bologna.