GROWTH AND MARKOV CHAINS : AN APPLICATION TO ITALIAN PROVINCES

SALVATORE PONZIO AND LUCA DI GENNARO

Abstract. This paper collects the main criticisms and limits present in literature linked with the use of distribution dynamics as a method for studying the evolution in time of an entire cross-section distribution. It then applies the suggested procedure to overcome each of them to per capita income data for 92 Italian provinces over the period 1952-1995. It then proposes, as a future line of research, the use of generalized Ehrenfest urn in order to overcome the criticism of the time homogeneity of Markov transitional Matrix, not explicitly solved in the empirical application. It is, in fact, a particular markovian stochastic process characterized by a time varying entries of the transitional matrix at each iteration.

Introduction

In the last few years an increasing body of literature has been focused in the Growth theory and in particular in the convergence of world income distribution. Substantial part of literature known as "new growth theory" has used cross sectional and panel data regression analysis focusing on the behavior of a representative economy and making the implicit assumption that each economy is characterized by a steady-state growth path along which the economy is moving (Baumol, 1986; Barro, 1991; Barro and Sala-i- Martin, 1991, 1992, 1995; Mankiw et al., 1992; Sala-i-Martin, 1996, amongst others)[1]. However, Quah points out that concentrating on the behavior of a representative economy can only shed light on the transition of this economy towards its own steady-state whilst giving no information on the dynamics of the entire cross-sectional distribution of income (Quah, 1993b, 1996a,b); that cross country income data do not seem to support the hypothesis of a smooth monotonic transition to the steady-state, showing instead a strong instability in the underlying process of growth (Quah, 1993a, 1996b); that the convergence rate found in many cross-sectional regressions could arise for reasons which are independent from the dynamics of economic growth; Quah (1993, 1997) has then proposed an innovative econometric method to describe the dynamics of income across countries suggesting to concentrate directly on entire cross-sectional distributions of per capita income, using stochastic kernels to describe their law of motion. Purpose of the analysis is to find the law of motion of this distribution, rather than simply computing few - less informative - moments of it. The implications for the convergence debate are then drawn either on the basis of the ergodic distribution of the process, in the discrete case, or directly analysing the shape of a three-dimensional plot of the stochastic kernel, in the continuous case.

Quah (1993a) classifies countries into groups by relative income, and estimates a transition matrix giving the probability that countries move between groups. Quah's application of the simple Markov chain model reveals evidence of a world in which the rich and the poor are diverging to form 'twin peaks'. This twin peaks result has motivated theoretical work on

I would like to thank Prof. Guido Pellegrini for the availability of data and Prof. Stefano Magrini for his helpful suggestions. I would especially like to thank Francesco Turino for support and realization of econometric program.

[1]The steady-state growth path doesn't need to be the same for all economies.

growth models with multiple steady states, in which countries above a cutoff level of income converge to a high income level, while those below the cutoff fall into a poverty trap. At a general level, this paper, in the spirit of the approach suggested by Quah, directly analyses the cross-sectional distribution of per capita income, studying its intra-distributional dynamics and the change in its external shape. Differently from Quah (1996a), the present analysis makes use of a discrete income space rather than a continuous one because discretisation not only allows the study of the one-period dynamics and the resulting ergodic distribution, but also the analysis of the transitional dynamics as well as the calculation of the speed at which the steady-state is approached. In what follows this non parametric approach is applied to study the evolution of regional disparities in per capita income within the 92[2] Italian provinces over the from 1952 to 1995. The rest of the paper is organized as follows. Section 1 describes the Markov Chains at a general level and their particular use in the study of economic convergence. In particular Subsection 1.2 collects the main criticism present in literature linked with the use of distribution dynamics as a method for studying the evolution in time of an entire cross-section distribution. Section 2 presents the empirical results obtained taking into account the suggested procedure to Italian provinces. In order to avoid annoying misunderstandings it is useful to remark the empirical nature of the present work. The theory behind it is not originally developed by the author but the standard theory that can be found in all books of stochastic process. Section 1 and Subsection 2.1 can be considered only a collection of other authors contributions. It has been reported only to give the reader an exhaustive but compact idea of the theory behind the empirical application.

Section 3 concludes describing the basic Ehrenfest urn model and proposing the use of generalized Ehrenfest urn in order to overcome the criticism of the time homogeneity of Markov transitional Matrix, not explicitly solved in the empirical application.

1. A Markov chain approach to the study of convergence

1.1. **General approach.** A (finite, first-order, discrete) Markov chain is a stochastic process such that the probability p_{ij} of a random variable X being in a state j at any point of time t+1 depends only on the state i it has been in at t, but not on states at previous points of time:

$$\begin{aligned} Pr\{X_{t+1} = j \backslash X(0) = i_0, ..., X_{t-1} = i_{t-1}, X(t) = i\} \\ = Pr\{X_{t+1} = j \backslash X_t = i\} \\ = p_{ij} \end{aligned}$$

If the process is constant over time the Markov chain is completely determined by the Markov transition matrix

$$\prod = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}, p_{ij} \geq 0, \sum_j p_{ij} = 1$$

[2]In more recent year some new provinces have been created, but after 1995 so it eliminate every possible problem in construction data.

which summarizes all N^2 transition probabilities p_{ij} $(i, j = 1, , N)$, and an initial distribution $\mathbf{h}_0 = (h_{10}h_{20}....h_{N0})$, $\sum_j h_{j0} = 1$, describing the starting probabilities of the various states.

The Markov chain approach to the study of convergence across economic systems adopted here can be described as follows. Denote regional per capita GDP at time t relative to the top five Italian provinces average by y^t and the corresponding cross sectional distribution by d^t [3] . Define a set C of m non-overlapping income classes on the basis of some criterion. The evolution of this distribution over time can be described by the following equation:

$$d^{t+1} = P \cdot d^t \tag{1}$$

where P describes the transition form one distribution to the other and $d^t = (d^t_1, ..., d^t_m)'$ is the vector of population proportions at time t. In other words, P can be interpreted as a transition probability matrix: for any two income classes i and j $(i, j \subset C)$, the elements define the probability of moving from class i to j between time t and *t+1*. Supposing that a province r is in class i $(y^t_r \in i)$ at time t, if the sequence $\{y^0_r, y^1_r, ...\}$ satisfies the relation

$$Pr\{y^{t+1}_r \in i \backslash y^t_r, y^{t-1}_r, ..., y^0_r\} = Pr\{y^{t+1}_r \in i \backslash y^t_r\} \tag{2}$$

for any $i \in C$, and for any region, then the evolution of the cross-sectional distribution d described by Eq. (1) can be analyzed as a time-homogeneous (finite)Markov chain [4]. If P is ergodic[5] , it is well known that the chain is characterized by a stationary distribution where the conditional probability of occupying an income class in the next period is the same as the unconditional probability. This stationary distribution is described by the following limiting probabilities:

$$\pi_j = \frac{A_{jj}(1)}{\sum_{kA_{kk}}(1)} j, k \in C \tag{3}$$

where $A_{jj}(1) = A_{jj}(\lambda_1)$ is the (j, j)th cofactor of the matrix $A(\lambda_1) = A(1) = (I - P)$. The existence of a stationary distribution can be investigated directly by considering the second eigenvalue of P. When the modulus of the second eigenvalue, λ_2, is strictly smaller than 1, the transition probability matrix converges to a limiting matrix P* and the cross-sectional distribution converges to its steady-state. In this case, the speed with which the steady-state is approached can then be evaluated resorting to the concept of asymptotic half life of the chain, *hl*, that is the amount of time taken to cover half the distance from the stationary distribution. The half life is defined as (Shorrocks, 1978):

---

[3] Regional per capita income is measured in relative terms to the five leading provinces. This makes it possible to separate the effects on the cross-sectional distribution of aggregate (Italian) forces from the effect derived from regional-specific forces, having conditioned their aggregate effects out.

[4] Note that the definition of Markov chain in Eq. (2) is not the standard one since it implies that the transition probability between any two states (income classes in the present case) is independent of time. This assumption of time homogeneity may appear particularly strong since economic conditions and policies change over time, implying changes in the transition probabilities. However, it should be noted that this assumption is equivalent to analyzing convergence towards the steady-state by running (cross-sectional or time series) regressions over a limited period of time. The general aim of all these approaches is to shed some light on the nature of the process of economic development that has characterized the Italian provinces during the time span covered by the data, and not to forecast what will happen in the future.

[5] A Markov chain is ergodic if it is irreducible and all its classes are ergodic. A Markov chain is said to be irreducible if the set of all classes C on which it is defined has no other closed subset other than C itself, where closed means that any class outside the set cannot be reached from any class inside it. A class is said to be ergotic if it is aperiodic and for which the expected average time of return to the class itself as finite. The typology of classes adopted in the text is standard and can be found in traditional textbooks (see, for instance, Chung, 1960; Kemeny and Snell, 1976).

(4) $$hl = \frac{-\log 2}{\log|\lambda_2|}$$

and ranges between infinity - when the second eigenvalue is equal to 1 and a stationary distribution does not exist - and 0 - whenλ_2 is equal to 0 and the system has already reached its stationary equilibrium. To sum up, the choice of the income grid, by providing a discrete approximation of the cross-sectional distributions under study, uniquely determines the m^2. transition probabilities that form the matrix P. The existence of a steady-state distribution is then investigated by studying the eigenvalues of this matrix; when this stationary distribution exists, the speed with which the system approaches it is calculated using the asymptotic half life of the chain in Eq. (4). For ergodic Markov chains, the stationary distribution can be computed directly using the limiting probabilities in Eq. (3)[6] [7]. Since the whole process is usually assumed to be time-invariant in the literature on income convergence, the transition matrix can be used to describe the evolution of the income distribution over any finite or infinite time horizon. The regional income distribution after m transition periods (from t to any $t+m$) can be calculated by simply multiplying the transition matrix m times by itself, using the income distribution at time t as a starting point, i.e. $h_{t+m} = h_t\Pi^m$. Moreover, if the Markov chain is regular the distribution converges towards a stationary income distribution h* which is independent of the initial income distribution h ($lim_{m\longrightarrow\infty}h^m = h^*$). Comparing the initial income distribution (h_0) to the stationary distribution (h*) is informative as to whether a system of regions converges or diverges in per-capita income. Higher frequencies in median-income classes of the stationary than the initial distribution indicate convergence, and higher frequencies in the lowest and highest classes indicate divergence.

Finally, important information about the dynamics of the cross-sectional distribution are provided by the higher order transition probabilities, $p_{ij}(l)$, which give the probability that, starting from income class i, a regional economy will enter income class j after exactly l periods, regardless of the number of entrances into j prior to l. In general, higher-order transition probabilities are characterized by the following relationship

(5) $$p_{ij}(l_1 + l_2) = \sum_k p_{ik}(l_1)p_{kj}(l_2)\forall i, j, k \in C$$

also known as the Chapman-Kolmogorov equation for time-homogeneous Markov chains. Through this relationship it is then possible to analyze the shape of the cross-sectional distribution d after any given number of time periods and hence to study directly the transient behaviour of the system. When the asymptotic half life of the chain is very short, the transient behavior might be of relative interest, but when the transition towards the steady-state is slow, the study of the transition towards the steady-state becomes very important, possibly more important than the steady-state itself.

[6]However, even if the transition probability matrix P is not ergodic, it still might be possible to derive a stationary distribution. In order to do this, it is necessary to identify which income classes are transient and which are ergodic. Since all transient classes will be empty when the system reaches its steady-state (Chung, 1960), the analysis of the stationary distribution can be carried out by identifying irreducible subsets of C, whose elements are all ergodic. Analogously to the general case, the speed of the transition phase of the system can be derived from the asymptotic half life of the individual ergodic subsets of C. Recall A class is called transient if there is a positive probability that starting from this class, a region will not return to the same class in a finite number of periods.

[7]It should be noted at this point that if the matrix P can be partitioned into k ergodic stochastic submatrices, the first k eigenvalues of P are all equal to 1, whilst the speed of the transition to the steady-state is governed by the (k+1)th eigenvalue.

1.2. **An important distinction: discrete and continuous processes.** Calling the variable of interest for the researcher at time t, Xt (with t an integer) and assuming it can take values in a certain set E. In the present framework, the variable of interest is the per capita income of the Italian provinces. Let F be the distribution of that variable at time t : describe its law of motion by a first order autoregressive process (Quah, 1997):

$$F_{t+1} = T^*(F_t) \tag{6}$$

where the operator T* maps the distribution from period t to period t+1. If Xt is discrete, that is it can assume only finite or countable number of values[8], the operator T* can be interpreted as the transition probability matrix M of a Markov process:

$$\phi_{t+1} = M'\phi_t \tag{7}$$

The elements $p_{x,y}$ of M are the probabilities of transition from state x to state y in one step. However, in many case - and for most economic variables - Xt can take infinite values, for example any number on the real line. In this case, the operator T* must be interpreted as a *transition function* or *stochastic kernel* $P(x, \cdot)$[9].

Let A be any subset of E; then the distribution at time t+1 is defined by

$$F_{t+1} = \int P(x, A)F_t(dy) \tag{8}$$

Thus the transition function maps the distribution Ft from one period to the other[10].

Although it assumes a markovian structure of the underlying process, the approach of distribution dynamics however, is different from the traditional Markov process theory. In the latter, the emphasis is on a scalar process, from which an unobservable sequence of probability distributions is usually inferred. Distribution dynamics shows his originality in the fact that a sequence of entire (empirical) cross section distributions is actually observed, while the (*dual*) scalar process is implied but never observed (Quah, 1996a). It is possible, however, to construct such (artificial) scalar process on the basis of some initial distribution and its law of motion described by the stochastic kernel, elements that completely characterize a Markov process.

1.3. **Discretization and other limits.** The stochastic kernel is a useful tool to analyze the dynamics of the entire distribution of the process. It could be useful, however, to "discretize" the state space, that is to partition the continuous state space in a finite number of intervals. The theory of finite state space Markov processes, in fact, is accessible and well developed . The estimation of the transition matrix is computationally simpler and results are easier to interpret and present; many indices and statistics are also easier to compute. Note that the issue of time discretization is very different from the one analyzed here, which is the discretization of the state space of the process.

The method of distribution dynamics was initially employed in a discrete framework (Quah, 1993): relative per capita incomes were divided into classes and transition probabilities between classes were estimated. The partition was either in equi-sized cells or in cells

[8] i.e. the set E is either finite or countably infinite.

[9] The stochastic transition function, or stochastic kernel, $P(x; A)$, describes the probability that the next step will take us in a certain set A, given that we are currently in state x:

$P(x; A) = Pr(\mathtt{X}t + 1 \in A \backslash Xt = x)$

for all values x in E and all the subsets A.

[10] Let p $(x; y)$ be a measurable function which is non-negative: $p(x;\ y) \geq 0$, and integrates to one:$\int p(x; y)dy = 1$ (where *x,* y *are points in E*). The kernel $P(x; A)$ can be defined as the integral of this function over the set A: $P(x; A) = \int p(x; y)dy$ then $p(x; y)$ is the *transition density function* associated with P $(x; A)$.

with variable upper endpoint, with approximately the same number of occurrences in each group (Quah, 1993, 1996b). To avoid the possible bias induced by the discretisation, Quah (1997) refines the analysis employing a continuous state space framework, but the discrete version remains very popular in the literature.

All the discretisation methods above share a common problem: arbitrary discretisation of the state space is almost certain to remove the Markov property of the process. Consider, for example, an equi-sized discretisation of the distribution of per capita incomes. Assume that the state space is partitioned in a finite numbers of intervals of the same size. But as has been shown, Bulli (2001), any arbitrary discretisation corresponds to creating a partition of the space into a finite number of subsets A1....Aj and then associating each subset with a distinct state in a discrete state space. Thus, in terms of the underlying scalar process, this is equivalent to creating a sequence:

$$\eta^{(t)} = \sum_j j I_{A_j}(X_t) \tag{9}$$

where I_{A_j} is the indicator function, so that $I_{A_j} = 1$ if $X_t \in A_j$, 0 otherwise .

The sequence η^t in (9) is not normally a Markov chain because of the dependence on the previous values of η^k. The consequences, from a theoretical point of view, is that the obtained matrix is not a "transition probability" matrix[11], as the properties of the underlying process are not clear. Nor calculating the limiting distribution for that process from the estimated transition matrix is correct[12]. The loss of the Markov property can have severe consequences for the analysis, Bulli (2001).

Another relevant problem in using these approach is due to the issue of filtering, Hites (2001). In the simple Markov chain model, the estimated transition probability matrix is used to extract information concerning the mobility of countries, or provinces as in this work, within the distribution of incomes. This information is camouflaged by two sources of noise. The first is generated by inaccuracies in the data and will not be discussed here. The second results from using continuous data to estimate a discrete model (i.e. from translating continuous data into discrete data by defining income class frontiers). Whereas the first source of noise affects the inference in ways that are unknown for us, the second source of noise can be observed directly. In the simple Markov chain model, transitions represent mobility. In reality, however, transitions can occur for two reasons. Transitions can result from higher (or lower) than average country's growth in a province (see Figure 1) and that is what we would like to measure. Transitions can also result from business cycle type variations in a region's income when the level of income is situated very close to one defining a frontier between classes (see Figure 2).

Clearly, this is not mobility and since such transitions are included in the calculation of mobility, it is necessary to purge the data of such noise. So, it needs to correct for the bias that short run fluctuations in income introduce into the calculation of long run tendencies in the distribution of country incomes[13]. Two remarks are in order. First, the noise introduced by business cycle type fluctuations at frontiers between income classes only affects estimates of the probability of transitions between adjacent states, that is estimates of the elements on the diagonals just above and below the main diagonal of the transition probability matrix. Given the tri-diagonal structure and the stochastic nature (i.e. all rows sum to one) of

[11]Transition probabilities are usually estimated as the sample proportion of the transitions from a given cell to another, relative to the total transitions from that cell.

[12]The limiting - or invariant - distribution π of a discrete Markov chain with transition matrix M satisfies: $\pi = M'\pi$. Thus π describes the long run, stable behavior of the process.

[13]This is an estimation problem arising from the bias introduced into the transition matrix by fitting a discrete model to a finite sample of continuous data. In theory, the ergodic distribution is independent of short run noise.

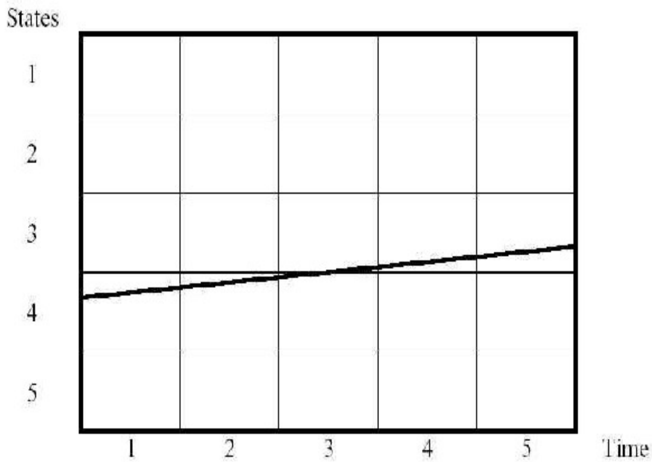

FIGURE 1. Transitions resulting from higher than average variations

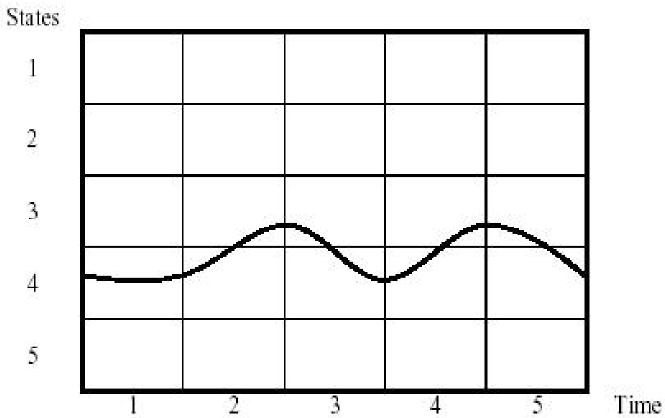

FIGURE 2. Transitions resulting from business cycle type variations

the estimated transition probability matrices, all the elements of the estimated transition probability matrices are potentially affected by this noise. In this case the elements on the main diagonal of the estimated transition probability matrix are underestimated, the rest of the elements overestimated, and therefore mobility is overestimated.

The transition matrix can be estimated by a Maximum Likelihood (ML) approach. Assume that there is only one transition period, with the initial distribution $h = n_i$ /n being given, and let n_{ij}denote the empirically observed absolute number of transitions from i to j. Then, maximizing

$$lnL = \sum_{i,j} n_{ij} lnp_{i,j}, s.t. \sum_{j} p_{ij} = 1, p_{ij} \geq 0$$

with respect to p_{ij} gives

$$\widehat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$$

as the asymptotically unbiased and normally distributed Maximum Likelihood estimator of p_{ij} [14]. The standard deviation of the estimators can be estimated as

$$\widehat{\sigma}_{p_{ij}} = (p_{ij}(1 - p_{ij})/n_i)^{1/2}$$

Obviously, the reliability of estimated transition probabilities depends on two aspects: First, the data-generating process must be Markovian, i.e. meet the assumptions of Markov chain theory (Markov property, time-invariance). Otherwise, the estimators $\widehat{p}_{ij}$ are not allowed to be interpreted as Markov transition probabilities, and cannot be used to derive a stationary distribution. And second, the estimates have to be based on a sufficiently large number of observations. Otherwise, the uncertainty of estimation is too high to allow for reliable inferences.

In practice, the estimation of Markov chains is subjected to the trade-off between increasing the number of observations to obtain reliable estimates, and increasing the probability of violating the Markov property. Given that data availability is limited in the geographic as well as in the time dimension it would be, in principle, preferable to estimate the probabilities from a data set pooled across time and space, using as many transition periods and regions as possible. With regard to the Markov property, however, the regions should not be too small. The smaller they are, the higher will be the intensity of interaction, and thus the correlation of income levels, between neighboring regions tends to be. On the other hand, by extending the geographical coverage of the sample,it will increase the danger of lumping together regions whose development patterns are heterogeneous. Single regions, or certain groups of regions may follow development paths which are different from other regions'paths.

Likewise, the longer is the time period under consideration, the higher will be the risk of structural breaks, i.e. regime changes which seriously affect the evolution of the income distribution[15]. As a consequence, the evolution prior to the shock may not be informative for the subsequent evolution of the income distribution; the stationary income distribution (h*) estimated from a transition matrix for the entire sample may be misleading.

Moreover, the estimated ergodic probabilities are very sensitive to changes in estimated transition probabilities. Therefore, even small revisions in the data may change substantially the shape of the estimated ergodic distribution .

2. An empirical application

2.1. General procedure. Crucial to the implementation of the Markov chain approach described above is the estimation of the transition probability matrix P. This, in turn, requires the estimation of the (unknown) probability density function that has generated the observed data. Given n provinces, the cross-sectional distribution of per capita income at time t, represented by the vector $(y_1^t, y_2^t ..., y_n^t)$, can be seen as a random sample from a continuous probability function f where

[14] This assumes that the initial distribution h does not contain any information about the Markov process and, thus, the transition probabilities p_{ij}.

[15] As Fingleton (1997) notes, the Markov chain approach is well suited to capture an uneven stream of small shocks that affect economies from time to time. Large, one-off shocks, however, are not consistent with time-invariance of transition probabilities.

$$f(y^t) \geq 0, \int_{\Re} f(y^t) dy^t = 1$$

So, by following this approach,all the analysis must provide an estimate of f[16] when no formal parametric structure is specified. One of the ways suggested by Quah (1993b), (1994) proceeds by using m quantiles, and then calculating the corresponding fractile transition probability matrix. In the discrete version of this approach, Quah (1993a) divides countries into five groups: those with less than 1/4 of the world average per capita income; those between 1/4 and 1/2 of world average income; those between 1/2 world average income and world average income; those between 1 and 2 times world average income, and those with income greater than twice the world average.

This work follows Quah [1993a] in assuming that each country's relative income follows a first-order Markov process with time-invariant transition probabilities. That is to say that, a country's (uncertain) income tomorrow, only depends on its income today.

But it differs from it since the income is measured relative to the income of five leading provinces [17].

In order to cleanse the data from the noise introduced by business cycle fluctuations described in subsection 1.3, it is necessary to tighten the conditions under which a transition is considered to represent mobility. Here this is achieved by requiring a transition to last a minimum number of periods so it can be counted as mobility, this minimum number of periods being defined as just over the average number of periods spanned by a business cycle. The HP filter filters is applied to the original data in order to eliminate cyclical component from series[18].

As described in subsection 1.3 the totally subjective choices may represent a source of potential problems since inappropriate discretisation can remove the Markov property from a first-order Markov process. Even though, as Quah points out (Quah, 1996b), the distortions introduced through discretisation are not likely to conceal the most important features of the distribution dynamics under study, it seems nevertheless important to reduce the existing degree of arbitrariness. The starting point is the recognition that the arbitrary discretising grid used to construct the transition probability matrix is in fact a crude nonparametric estimator of the probability density function that generated the observed cross sectional data. It is therefore interesting to make use of other nonparametric methods that estimate a density function. Thus, in the present case attention is concentrated on the histogram with equisized cells , a consistent estimator of the true underlying probability density function that, at the same time, provides a discrete approximation of the continuous cross-sectional distribution. In an attempt to reduce the degree of arbitrariness in order to avoid the possible bias then induced , this work follows Magrini (1999) suggesting a procedure that

---

[16] In other words, f is taken to belong to a large enough family of density functions so that it cannot be represented through a finite number of parameters. The resulting estimate of f is therefore a nonparametric estimate.

[17] Quah's procedure, in fact, can potentially generate a long-run distribution of world income, in which all countries have above average income. Classifying countries' income relative to that of the leading countries rather than relative to world income eliminates this anomaly. In independent work, Pearlman [2000] also notes the possibility that a transition matrix estimated with Quah's income groupings can lead to a logically impossible ergodic distribution of income. His approach to resolving this problem involves categorizing countries relative to the geometric mean of income among countries of the world.

[18] The Hodrick-Prescott filter involves, in essence, defining cyclical output y_t^c as current output y_t less a measure of trend output y_t^g, with trend output being a weighted average of past, current and future observations: $y_t^c = y^t - y^g = y_t - \sum_j a_j y_{t-j}$. The HP filter is then derived by solving the following minimization problem : $min_{\{y_t^g\}_{t \longrightarrow \infty}} \sum_{t \longrightarrow \infty} \{(y_t - y_t^g)^2 + \lambda[(y_{t+1}^g - y_t^g) - (y_t^g - y_{t-1}^g)]^2\}$. For annual data, the standard value chosen for the smoothing parameter λ is 400. When $\lambda = \infty$ the solution to this problem is a linear trend, while with $\lambda = 0$ the trend coincides with original series.

eliminates subjectivity in the choice of the income class size, by concentrating on histograms as approximations to continuous distributions of income across regions. The method consists in choosing the bin width h for the histogram optimally, such that the same grid allows to discretize the empirical distribution at two different points in time, so that any measure of the error of approximation (mean-squared error or integrated absolute error) is minimized. The choice between possible grids of income classes is thus made in terms of the ability of the discrete distributions to approximate the observed (continuous) distributions[19] . The choice of the other parameter h is, however, quite important. If h is too small, then the histogram will be too rough and it will become more likely to find closed subsets of the transition probability matrix P that do not communicate; on the other hand, if h is too large, then the histogram will be too smooth, resulting in the loss of important information on intra-distributional dynamics. The choice of h should therefore be balanced between these two extremes by minimizing a measure of the error of approximation.

Assuming that the probability density function that has generated the observed distributions is square integrable a possible approach generally measures the performance of the estimator of the probability density function, $\widehat{f}$, in terms of the integrated mean squared error (IMSE); following this approach, Scott (1979) shows that the optimal bin width, h , can be estimated by

$$\widehat{h}_n^* = 3.49sn^{-\frac{1}{3}}$$

where s is the sample standard deviation. An alternative criterion within the same approach has been suggested by Freedman and Diaconis (1981). Using the interquantile range, IQR, of the data as a measure of the scale of the random variable under study, they suggest a simple data-based rule

$$\widehat{h}_n^* = 2(IQR)n^{-\frac{1}{3}}$$

As shown by Devroye and Gyorfi (1985), one problem with this approach to nonparametric density estimation is that the tail behaviour of a density becomes less important, possibly resulting in peculiarities in the tails of the density estimate. To overcome this problem, which is particularly relevant to the type of analysis pursued here, these authors develop an alternative approach that focuses on the integrated absolute error (IAE). When f is assumed to be a normal distribution $N(0, \sigma^2)$, they obtain

$$\widehat{h}_n^* = 2.72sn^{-\frac{1}{3}}$$

that is shown to perform quite well also for non-Gaussian samples.

The procedure proposed here is therefore represented by the calculation of a set of estimates of the optimal income class size by applying different optimizing criterion to each of the observed distributions . The value thus calculated are then applied to the observed vectors of income data in order to derive discrete approximation of the observed distribution based on a common grid of income classes. Then comes the choice between the different grids in terms of the ability of the corresponding discrete distributions to approximate the observed distributions. As a consequence, all analyzes start with a normality test[20]on the available data. Finally, Quah (1993a) estimated a transition matrix using annual data on GDP per capita, it seems natural however, as suggested by Kremer, Onatsky and Stock (2001), to consider transition periods longer than 1 year. The assumption of a one-period

[19]The intervals have been chosen closed on the left and open on the right for definiteness.

[20]It is important to note that although the criteria for the optimal choice of the income class size are derived under the assumption of normality for the unknown probability density function, they are generally able to deal with non-normal samples provided that the deviations from normality are not too great

Markov process is likely to be violated. A group 3 country that experiences a recession in a particular year and falls just over the borderline into group 2 is less likely than other group 2 countries to fall into group 1 the next year, and more likely than most group 2 countries to transit to group 3 in the following year. Considering transition periods longer than one year reduces the impact on the estimated transition matrix of high frequency fluctuations in income of countries that happened to be close to the threshold between different groups at the beginning of the period. So in this work 4-year data are used, in the believe that using 4-year data may provide a more accurate picture of the long-run dynamics then using annual data.

2.2. **Empirical result.** The analysis is based on the updating of a newly compiled database[21] on the gross product and population of Italian provinces (corresponding to the NUTS 3 level in the official UE classification)[22].

A formal test of normality can be performed using the Skewness and Kurtosis[23] and the Shapiro-Wilk tests. Table 1 reports the tests statistics and the corresponding p-levels. It is clear that the normality assumption can be rejected for all period distributions under analysis. The procedure proposed here is, therefore, represented by calculation of a set of estimates of optimal income class size by applying the optimizing criterion proposed by Devroye and Gyorfi (1985) since it has been shown to be better performed than the other for non-Gaussian samples[24].

TABLE 1. Skewness/Kurtosis tests for Normality and Shapiro-Wilk test for normal data.

Distribution	Sk Test		S-W Test	
	adj χ^2	$Prob > \chi^2$	z	$Prob > z$
1952	12.09	0.0024	4.209	0.00001
1972	9.87	0.0072	1.950	0.02561
1995	9.95	0.0069	2.506	0.00611

It should be noticed that different results are obtained by comparing four years transition, table 2, and longer period transition, table 3. In the former probabilities are strongly concentrated on diagonal. The diagonal represents the probability of a region remaining in its original income group. The off-diagonals show the probabilities of regions to move into

[21]The development analysis of Italian provinces is based on reconstruction of a data set on provincial gdp per capita for the period 1952-1995. Data are obtained by annual Istituto Tagliacarne's estimation.

[22]The territorial basis units are administrative units defined for every UE country. A part NUTS 3 there are NUTS 0 (countries), NUTS 1 (regions for Italy), NUTS 4 (not defined for Italy) and NUTS 5 (towns).

[23]Sk test presents a test for normality based on skewness and another based on kurtosis and then combines the two tests into an overall test statistic. Sk test requires a minimum of 8 observations to make its calculations.

[24]More precisely the DG criterion applied to 1952 is chosen as the one providing the best approximations.

Table 2. Estimated transition matrix for 92 Italian provinces 1952-1995, 4 years transition.

	Final year					
Initial year	state 1	state 2	state 3	state 4	state 5	state 6
state 1	0.87					0.13
state 2		0.80	0.20			
state 3			0.84	0.16		
state 4			0.02	0.78	0.20	
state 5				0.03	0.82	0.15
state 6	0.06				0.05	0.89
Ergodic	0.25	0	0.003	0.027	0.18	0.54

Table 3. Estimated transition matrix for 92 Italian provinces 1952-1995, 44 years transition.

	final year					
initial year	state 1	state 2	state 3	state 4	state 5	state 6
state 1	0.34	0.24	0.24	0.17	0.01	
state 2		0.09	0.36	0.46		0.09
state 3			0.60	0.40		
state 4			0.25	0.75		
state 5				1.00		
state 6					1.00	
Ergodic	0	0	0.38	0.62	0	0

other income groups. Obviously for shorter period of time there is less probability to jump from a class (level of income)to another one.

More interesting is the comparison between the 44-year matrix transition from 1952 to 1995 obtained in this work, table 3, and that one obtained by Fabiani and Pellegrini (1997) in a similar analysis, table 4. Although following quite different methodology[25]matrices are very similar in their meaning.

Table 4. Fabiani and Pellegrini transition matrix, 1952-1995.

	Final year						
Initial year	state 1	state 2	state 3	state 4	state 5	state 6	state 7
State 1	0.40	0.24	0.32	0.04			
State 2	0.09	0.45	0.09	0.22	0.09		
State 3	0.08	0.08	0.23	0.15	0.08	0.20	0.15
State 4				0.09	0.27	0.36	0.27
State 5			0.22	0.11	0.33	0.33	
State 6			0.20		0.40	0.20	0.20
State 7			0.06	0.11	0.11	0.39	0.33
Ergotic	0.03	0.04	0.17	0.09	0.25	0.25	0.16

[25]In Fabiani and Pellegrini (1997), as in part of literature, provincial GDP per capita relative to the country has been discretised into 7 intervals so that the observed sample is approximately divided in equal categories. Moreover, are considered annual instead of four year period transition for data without applying any kind of filter.

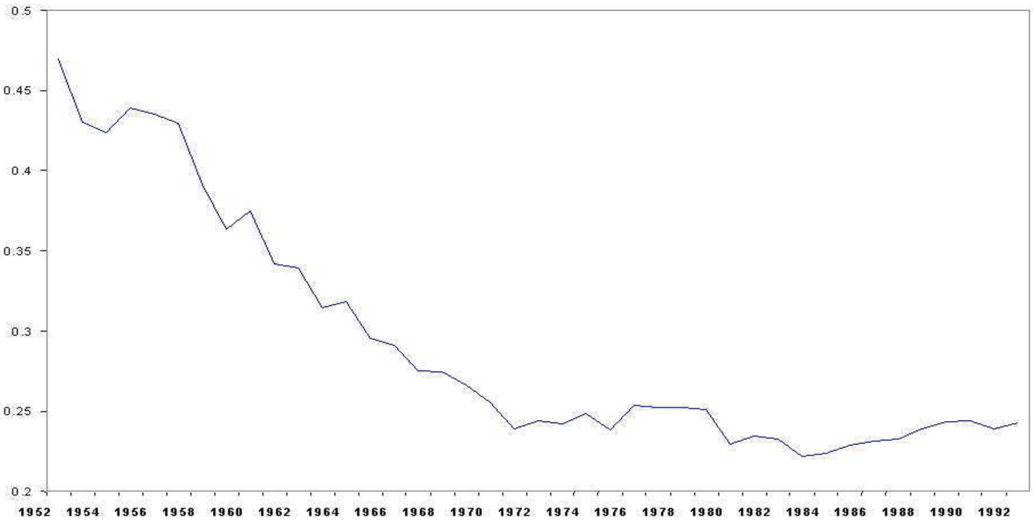

FIGURE 3. Normalized standard deviation of per capita GDP of italian provinces: 1952-1995.

Both tables show that mobility was high[26], especially for the poorest and richest groups. The last lines of these matrices show long run ergodic distribution. Again, considerable mobility was revealed but convergence was not so evident as the majority of provinces doesn't cluster in the middle income groups (Table 3 and 4). A concentration in intermediate states exists but not in the middle, leading to a bimodal distribution (twin peaks distribution). To go deep into details Table 3 shows maximum of mobility in the upper income classes where probability remaining in its original income group is zero while in Table 4 this feature is less evident even if probability to remain in the same income group are less then probability to change group[27]. That also happens because provinces that have shown most propensity to leave their class are those near biggest industrialized areas (above all north-east provinces of Treviso, Vicenza, Forli' integrated by some central provinces as Pesaro, Urbino, Arezzo and Lucca and some very dynamic provinces in the south of Italy as Bari, Taranto, Avellino,Palermo, Catania).

Moreover from the analysis of variance provincial GDP per capita, see figure 3, it can be noticed a clear stylized fact : variance has been reduced of more than half from the beginning to the end of the period and the reduction has been concentrated in the period 1952-1970.

The process of convergence is evident till '70 when it has been stopped and has inverted its tendency during '80. This feature is even more clear with Markov chain approach. The whole period has been then divided into two subperiods : 1952-1972 and 1973-1995. As shown in table 5, the 1952-1972 subperiod transitional matrix shows a very huge process of convergence (for all classes probability to remain in the same class in the successive period is very low compared with probability of moving). This tendency is even more evident

[26]The Shorrocks(1978) mobility index given by $m = \frac{n-Tr(\Pi)}{n-1}$ where Π is the transitional Matrix, $Tr(\cdot)$ the trace and n number of class and ranging between 0 and 1, for the period 1952-1995 was m = 0.844.

[27]In the present work provincial income per capita is weighted to top five leading provinces so that transition towards these provinces are more sensitive.

looking at ergotic distribution wich tends to be unimodal. More interesting considering the second eigenvalues of the matrix of interest can be determined the speed of the transition towards the stationary distribution. The second eigenvalue for this matrix is equal to 0.40 leading to half, life mentioned in subsection 1.1, of 0.75647 time periods. This means that the whole system needs more or less 4,5 years ($0.75647 \times 6 = 4.53882$) to reach half a way to its steady state. If the convergence process was continued following that tendency for other ten years the poorer provinces should get the income per capita level of richest provinces and the convergence should be very strong. Table 6 shows that the opposite holds for the 1973-1995 subperiod transitional matrix and ergotic distribution. Also for these cases the results are very similar to those produced by Fabiani and Pellegrini (1997) (Table 7).

Table 5. Estimated transition matrix for 92 Italian provinces 1952-1972, 4 years transition.

			Final year			
Initial year	state 1	state 2	state 3	state 4	state 5	state 6
state 1	0.36	0.19	0.35	0.10		
state 2			0.18	0.82		
state 3			0.40	0.60		
state 4				0.25	0.75	
state 5				0.33	0.67	
state 6					1.00	
Ergodic	0	0	0	0.31	0.69	0

Table 6. Estimated transition matrix for 92 Italian provinces 1973-1995, 4 years transition.

			Final year			
Initial year	state 1	state 2	state 3	state 4	state 5	state 6
state 1	0.20	0.80				
state 2	0.14	0.72	0.14			
state 3		0.24	0.38	0.38		
state 4		0.13	0.25	0.19	0.30	0.13
state 5			0.05	0.25	0.35	0.35
state 6				0.08	0.38	0.54
Ergodic	0.06	0.37	0.15	0.13	0.14	0.15

TABLE 7. Fabiani and Pellegrini ergotic distribution.

Period	State 1	State 2	State 3	State 4	State 5	State 6	State 7
1952-1970	0	0	0	0	0.48	0.52	0
1970-1995	0.26	0.13	0.15	0.13	0.10	0.11	0.11

Italian economic development has followed, in the period under consideration, different patterns for different periods: in the first part , after the second world war, Italy experienced the so called 'economic miracle', characterized by high economic growth rate (average growth rate is equal to 5,4 per cent, higher then average growth rate of the century) all over the country. Such a fast growth led to the saturation of the already industrialized areas and decentralization toward new developing provinces . This period is also characterized by an increasing government expenditure (from 9.5 to 16.7 per cent of GDP) towards regions less developed. Under public incentive, the first significant nuclei of modern industrialization grew up in the southern provinces. This occurrence could be seen as the main reason for the strong process of convergence observed during the period took into consideration. A second phase starting from 70' has been characterized by a high degree of opening of the Italian economy to international markets. This increased the impact of exogenous shocks and international business cycle. Southern provinces due to their low degree of opening didn't suffer the effects of international crisis continuing the convergence process but, on the other hand, they didn't have the opportunity to participate to successive benefits coming from international trade. Finally from '80 convergence process begun in earliest '50 has shown an huge decrease till an inversion of tendency in the last period under consideration. Last years have continued to follow this tendency. In 2000 in a contest of acceleration of economic performance there was an evident divergence between different provinces. In conclusion one of the main Italian economic development characteristic is the deep territorial difference[28]. The results obtained by policies for Mezzogiorno[29] development in the last fifty years lead to think about Italian case as a clear example of dualistic growth in the same time of a fast integration between regions. In the period under consideration the process of convergence has been significative but not enough to eliminate dualism present in the Italian economy.

3. The Ehrenfest urn model.

3.1. **The basic Eherenfest urn model.**

The Eherenfest urn model[30] , Eherenfest(1907),in its simplest version can be introduced as follows. There are two urns in which are distributed k balls. In each instant of time a ball$n$is drawn between the $k$balls from an urn and moved in the other urn[31]. Considering as state of the system the number of k balls in the first urn, there are k+1 states 0,1,2,3...,k.

When the system is in the state i,that is there are i balls in the first urn, the probability that the drawn ball is in the fist urn is $\frac{i}{k}$ that is also the probability that the system goes

[28]In the analysis of economic development of Italian provinces and areas it could be observed that provinces denoting better results are those characterized by a concentration of small and average firms belonging to a specific productive sector, the industrial one. This consists of a set of small firms specialized in different phases of the same productive process,with a series of local institution favorable to the integration, competitive and cooperative

[29]It indicates regions and provinces of the south of Italy.

[30]The Eherenfest's urn was first used in physics, in particular in statistical mechanism. It has subsequently seen a widespread application in economics, genetic, philosophy.. See for instance, Costantini Garibaldi (2004) page 516 and Costantini(2004), chapter 9, in particular page 224. In all book of stochastic process or probability the Eherenfest urn model can be found in the chapter of Markov chains. See for instance Feller (1950), page 377.

[31]There is in literature a funny introduction of Eherenfest urn model that consider in state of urn dog and in state of balls fleas.

to the state i-1. The probability that the system goes to the state i+1 is then $\frac{k-i}{k}$; This feature is due to the fact that the random chosen ball is forced to change urn. The other transition probability are zero:

$$p_{i,i-1} = \frac{i}{k} \qquad p_{i,i+1} = \frac{k-1}{k} \quad i = 0, 1, ..., k \tag{10}$$

So characterized the markov chain is irreducible[32] and periodic(of period 2).

Even if its basic version seems simple the Ehrenfest urn model is a powerful and ductile tool.

We propose the use of generalized Ehrenfest urn in order to overcome one of the main criticism underlined in literature for the use of Markov chain in the convergence literature, that is the time homogeneity of Markov transitional Matrix. As explained in subsection 1.1 from the transition probability matrix P it is possible to derive its stationary distribution where the conditional probability of occupying an income class in the next period is the same as the unconditional probability. But,as underlined in the subsection 1.3 the data-generating process and the markov transitional matrix should be time invariant otherwise it is not allowed to use the transition probabilities to derive the ergotic distribution.

For that reason we propose to use the Ehrenfest urn, a particular markovian stochastic process characterized by a time varying entries of the transitional matrix at each iteration. Once obtained the transitional matrix from real data as suggested above it is possible to obtain another matrix iterating the previous one allowing the transitional probabilities to vary instead of calculating the ergotic distribution. From the final matrix it will be possible to derive conclusions about the convergence of income pro capita of different provinces. For instance, it is well known that a diagonal matrix and so also a matrix with all entries equal will lead to an uniform ergotic distribution. In economic terms it means respectively that if each province remains almost for sure (with probability close to one)in its original income class or the probability of moving in each of the other class (including the original one)is the same, the income per capita of different provinces does not converge.

In order to derive this particular final matrix we need to consider two generalizations of the basic Ehrenfest urn model. First we consider more than two urns. In particular we consider N^2 urns, where N is the number of income classes (see subsection 1.1 page 2), assembled in N groups. At each iteration N balls are drawn, one for each group and each ball can be transferred only in the urns of the same group. Is important to note that the number of balls in each urn and each group reflects the entries of probability matrix.

Second we do not force the random chosen ball to change urn so that at any steps the drawn ball has a positive, (not strictly equal to zero) probability to be put back into the initial urn.

This last paragraph represents the driving line for our feature research.

Conclusion

The aim of this work was to analyze the Markov chain approach to the study of convergence across economic system. In particular some desirable features for convergence analysis have been underlined, such as the possibility to determinate a stationary income distribution directly modelling the dynamics of the evolving cross section distribution. But they have been also mentioned the limits and restriction on which this method is based. As an empirical exercise, it was proposed an application to Italian provinces income distribution over the period 1952-1995. The empirical procedure adopted has tried to take into account the main limits objective of criticism, namely the arbitrary discretisation of continuous state

[32]The Markov chain is said to be irreducible if in its transition matrix it is possible to go from each state to all other states.

space Markov process, the possibility of including business cycle fluctuations. So particular emphasis has been placed on the development of a procedure that reduced subjectivity in the discretisation of income space; a fine filter has been applied on four year data in order to reduce the impact on the estimated transition matrix of high frequency fluctuations in income of provinces that happened to be close to the threshold between different groups at the beginning of the period. The result of the analysis based on data set for 95 Italian provinces suggested that the process of economic growth at work in Italy during 1952-1995 has followed different patterns, reaching, at the end, a tendency towards divergence. Although the estimated ergodic probabilities are very sensitive to changes in estimated transition probabilities the results obtained are quite similar to those reached by other authors in previous works following a different methodology. The huge change on tendency occurred in the period under consideration confirms that the hypothesis of time invariance of transitional matrix is quite strong. But this justifies only partially criticisms from part of literature since the ergotic distribution, as emphasized by Quah (1993), and the steady state distributions should not be read as forecasts of what will happen in the future; it should be rather interpreted simply as a characterization of tendencies in the period under consideration, as a way to make more evident intra-distributional dynamics characterizing it.

Appendix

Table 8. Provinces by classes

Province	Class in 1952	Class in 1995	Province	Class in 1952	Class in 1995
AGRIGENTO	1	1	MESSINA	1	1
ALESSANDRIA	2	3	MILANO	6	5
ANCONA	1	4	MODENA	1	5
AOSTA	5	4	NAPOLI	1	1
AQUILA	1	2	NOVARA	3	3
AREZZO	1	3	NUORO	1	1
ASCOLI	1	3	PADOVA	1	4
ASTI	1	3	PALERMO	1	1
AVELLINO	1	1	PARMA	2	4
BARI	1	2	PAVIA	3	3
BELLUNO	1	3	PERUGIA	1	3
BENEVENTO	1	1	PESARO	1	3
BERGAMO	1	4	PESCARA	1	2
BOLOGNA	2	6	PIACENZA	2	4
BOLZANO	3	4	PISA	1	3
BRESCIA	1	4	PISTOIA	1	3
BRINDISI	1	1	POTENZA	1	1
CAGLIARI	1	2	RAGUSA	1	2
CALTANISSETTA	1	1	RAVENNA	2	4
CAMPOBASSO	1	2	REGGIOC	1	1
CASERTA	1	1	REGGIOE	1	4
CATANIA	1	1	RIETI	1	2
CATANZARO	1	1	ROMA	3	4
CHIETI	1	2	ROVIGO	1	3
COMO	2	3	SALERNO	1	1
COSENZA	1	1	SASSARI	1	2
CREMONA	1	3	SAVONA	3	3
CUNEO	1	3	SIENA	1	3
ENNA	1	1	SIRACUSA	1	1
FERRARA	2	3	SONDRIO	1	3
FIRENZE	2	4	SPEZIA	1	4
FOGGIA	1	1	TARANTO	1	1
FORLI	1	3	TERAMO	1	2
FROSINONE	1	2	TERNI	1	2
GENOVA	4	4	TORINO	5	4
GORIZIA	2	4	TRAPANI	1	1
GROSSETO	1	2	TRENTO	1	4
IMPERIA	4	3	TREVISO	1	4
LATINA	1	2	TRIESTE	4	4
LECCE	1	1	UDINE	1	4
LIVORNO	2	2	VARESE	4	4
LUCCA	1	3	VENEZIA	2	3
MACERATA	1	3	VERCELLI	5	4
MANTOVA	1	4	VERONA	1	4
MASSA	1	2	VICENZA	1	4
MATERA	1	1	VITERBO	1	2

Table 9. Number of Provinces in different years

Classes	**Number of Provinces in 1952**	**Number of Provinces in 1995**
1	68	23
2	11	17
3	5	24
4	4	25
5	3	2
6	1	1

References

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

Barro, R.J., Sala-i-Martin, X., 1991. Convergence Across States and Regions. *Brooking Papers Economic Activity*, 1, 107-182.

Barro, R.J., Sala-i-Martin, X., 1992. Convergence. *Journal of Political Economy*, 100, 223-251.

Barro, R.J., Sala-i-Martin, X., 1995. *Economic Growth.* McGraw Hill,London.

Baumol, W.J., 1986. Productivity Growth, Convergence, and Welfare:What the Long-Run Data Show. *American Economic Review*, 76,1072-1085.

Bickenbach, F., Bode E.,2001. Markov or not Markov - This Should Be a Question. *Kiel Institute of World Economics*, Germany.

Bulli, S., 2001. Distribution Dynamics and Cross-Country Convergence: Does Discretisation Matter? Approach. *Scottish Journal of Political Economy*, 48, 2, 226-243.

Chung, K.L., 1960. *Markov Chains with Stationary Transition Probabilities.* Springer-Verlag, Berlin.

Costantini, Domenico (2004), *I fondamnti storico-filosofici delle discipine statistico-probabilistiche*, Bollati Boringhieri.

Costantini, Domenico and Ubaldo Garibaldi (2004),The Ehrenfest fleas: from model to theory, *Synthese*, 139 (3): 513-514.

Dall'Aglio, G. (2000), *Calcolo delle probabilit*, Zanichelli.

Eherenfest, P. and T. Eherenfest, (1907), Uber zwei bekannte Einwande gegen das Boltzmannsche H Therem, *Phys Zeitschrift*, 8,311-316.

Fabiani,A. Pellegrini, G., 1997. Education, Infrastructures, Geography and Growth: an Empirical Analysis of the Development of Italian Provinces. *Temi di discussione*, Banca d'Italia, 323, November 1997.

Feller, William (1950), *An Introduction to Probability Theory and Its Applications*, vol.I, JOHN WILEY and SONS,New York.

Fingleton, B., 1997. Specification and Testing of Markov Chain Models: An Application to Convergence in the European Union. *Oxford Bulletin of Economics and Statistics*, 59, 3, 385-403, 27.

Garribaldi, U. e Penco, M.A. (2000), Eherenfest's Urn Model Generalized: an Exact Approach for Market Partipation Models, *Statistica Applicata*, Vol.12, n. 2.

Kemeny, J.C., Snell, L.J., 1976. *Finite Markov Chains*, Springer-Verlag, New York.

Kremer, M., Onatsky, A., Stock, J.. Searching for Prosperity. *NBER working paper*, 8250.

Magrini, S., 1999. The Evolution of Income Disparities Among theRegions of the European Union. *Regional Science and Urban Economics*, 29, 2, 257- 281.

Mankiw, G.N., Romer, D., Weil, D.N., 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107, 407-437.

Pearlman, Joseph G. 2000. Twin Peaks - A Reassessment, Economics Department, *London Guildhall University, Manuscript.*

Quah, D., 1993. Empirical Cross-Section Dynamics in Economic Growth. *European Economic Review*, 37, 426-434.

Quah, D., 1993. Galton's Fallacy and Tests of the Convergence Hypothesis. *Scandinavian Journal of Economics*, 95, 427-443.

Quah, D., 1994. One Business Cycle and One Trend from Many Disaggregate. *European Economic Review*, 38, 605-613.

Quah, D., 1996. Convergence Empiric Across Economies with (Some) Capital Mobility, *Journal of Economic Growth*, 1, 95-124.

Quah, D., 1996. Empiric 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 Economic Growth*, 2, 1, 27-59.

Rey, S., 2001a. Spatial Dependence in the Evolution of Regional Income Distributions. *Paper presented at the 40th Annual Meeting of the WRSA, Palm Springs, CA*.

Rey, S., 2001b, Spatial Empiric for Economic Growth and Convergence. *Geographical Analysis*, 33, 3, 195-214.

Sala-i-Martin, X., 1996. Regional Cohesion: Evidence and Theories of Regional Growth and Convergence. *European Economic Review*, 40, 1325-1352.

Shorrocks, A.F., 1978. The Measurement of Mobility. *Econometrica*, 46, 1013-1024.

Istituto Studi Economici, Universita' "Parthenope", Napoli

Dipartimento di Scienze Statistiche, Universita' di Bologna

Graduate Program in Economics and Management,Universitat Pompeu Fabra, Barcelona
E-mail address: salvatore.ponzio@upf.edu