International Monetary Fund

From the SelectedWorks of Rodrigo Garcia-Verdu

May, 2005

Income, Mortality and Literacy Distribution Dynamics across States in Mexico: 1940-2000

Rodrigo Garcia-Verdu

Income, Mortality, and Literacy Distribution Dynamics Across States in Mexico: 1940-2000[*]

Rodrigo García-Verdú[**]
Banco de México

This paper analyzes the dynamics of the distributions of per capita Gross Domestic Product (GDP), the infant mortality rate, and the adult literacy rate across states in Mexico between 1994 and 2000. It analyzes the hypothesis of convergence to a common level in these three indicators. The methodology used is the estimation of transition matrices and kernel densities for each of these indicators. The results show there is high persistence over time in the positions states occupy in the distribution of these welfare indicators, and suggest there is convergence to a common level of adult literacy, but not to a common level of per capita GDP or infant mortality.

1. Introduction

One of the main implications of the neoclassical growth model is the convergence of per capita income to a common level among economies with the same preferences, technology and population parameters. According to this model, poorer states, regions or countries will tend to grow faster than richer ones, eventually catching up with them.[1]

[**]E-mail: rgarciav@banxico.org.mx

[1]For a formal derivation of this implication of the neoclassical growth model, see Barro and Sala-i-Martin (2004), Chapters 2 and 3.

This implication is known in the literature as absolute b-convergence, and it has been one of the most widely tested hypothesis in all of economics. Tests of the convergence hypothesis have been performed using data from a cross-section of countries as well as from states and regions within countries.[2]

There are reasons to expect that convergence would occur at a higher rate across states and regions within a country than between countries, since there is higher mobility of labor and capital within a country than across countries. Moreover, the technology, preferences and population parameters may be more similar among states and regions within a country than across countries.

Nevertheless, income disparities within countries are often as large or larger than income disparities across countries. For example, the ratio of average per capita GDP over the 1995-2000 period between the richest state in Mexico, the Federal District, and the poorest state, Oaxaca, was nearly six.[3] This is greater than the ratio of average per capita GDP between the United States and Mexico over the same period, which was almost four.[4]

Furthermore, in some cases the income gaps across states do not seem to be narrowing over time. In the case of Mexico, for example, the states with the highest and lowest levels of per capita GDP in 1960 were also the Federal District and Oaxaca, respectively, and the ratio of their per capita GDP was also approximately six.

Thus, although there have been significant improvements in the national average, with per capita GDP more than doubling between 1960 and 2000, the gap between the poorest and the richest states has remained nearly constant during this period.[5]

While cross-country studies report conflicting evidence on the hypothesis of absolute b-convergence in per capita income depending on the sample of countries chosen, the time period analyzed, and the methodology employed,[6]

[2]The terms absolute and b-convergence are used to distinguish them, respectively, from the related concepts of relative and s-convergence. Absolute and relative convergence refer, respectively, to whether economies converge to the same steady state or to different steady states. For an explanation of the relation between the concepts of b and s convergence, see Barro and Sala-i-Martin (2004), Chapter 11.

[3]Mexico is composed of 31 states and a Federal District. The Federal District is not actually considered a state. Thus, one should refer to the 32 units as federal entities. For simplicity, we refer to them as states rather than federal entities throughout this paper.

[4]The figures for Mexico come from the state GDP data from INEGI described below. The cross-county data are taken from the World Penn Table Version 6.1, by Heston, *et al.* (2002).

[5]This fact stands in sharp contrast with the experience of other countries, including the United States. For example, in 1900 the ratio of per capita income of the richest state (Montana) to the poorest (North Carolina) was approximately 5.5. By 2000, the ratio of per capita income between the richest state (Connecticut) and the poorest (Mississippi) was approximately 1.95. See Barro and Sala-i-Martin (2004), Chapter 11.

[6]See Barro (1997), Jones (1997), Parente and Prescott (1993), Pritchett (1997) and Quah (1993).

most studies across states and regions find strong evidence in support of this hypothesis.[7]

The remarkable similarity across different country studies in the estimated rate at which their respective states and regions converge even led to the so called "iron law of economic convergence".[8]

Are states in Mexico converging to a common per capita GDP level —as predicted by the neoclassical growth model— or will some states be trapped in a low per capita income level while others remain in a high per capita income level? Are the poor southern states of Chiapas, Guerrero, and Oaxaca condemned to live in poverty, or can they eventually catch up with the more developed north?

These are some of the questions addressed in this paper. The answers to them have far-reaching implications for the design of public policies aimed at reducing regional inequalities.

This paper contributes to the study of economic growth and spatial inequality in Mexico by analyzing the evolution over time of the distribution of per capita GDP across states over the period 1940-2000. Furthermore, it analyzes the evolution of the distributions of the infant mortality rate and the adult literacy rate. As argued below, these are important welfare indicators whose distribution dynamics have not been hitherto studied. The paper also analyzes the hypothesis of convergence to a common level of per capita GDP, infant mortality and adult literacy among Mexican states.

As such, this paper is part of a growing body of research on economic convergence across states and regions in Mexico. This literature includes contributions by Aroca, Bosch and Maloney (2003), Caraza (1993), Cermeño (2001), Chiquiar (2005), Esquivel (1999), Juan-Ramon and Rivera-Batiz (1996), Messmacher (2000), and Navarrete (1995), among others. Most of these studies have analyzed convergence to a common per capita GDP level, and have used the cross-section regression methodology popularized by Barro (1991), and Barro and Sala-i-Martin (1991, 1992).[9]

While these studies of regional growth in Mexico use different data sets and methodologies, almost all of them present evidence in support of the absolute b-convergence hypothesis for the pre-1980's period, and show that convergence stopped or at least slowed down sometime during the mid 1980's.

This paper differs from the previous studies in two ways. First, it considers the issue of convergence to a common standard of living, rather than just convergence to a common level of per capita GDP. This has several potential advantages: (i) by focusing on a more comprehensive measure of development, it

---

[7]See, for example, Barro and Sala-i-Martin (1991, 1992, 2004).

[8]It should be noted, however, that most of the earlier evidence on regional convergence comes from studies of countries which are currently among the group with the highest per capita GDP, including Belgium, Canada, Denmark, France, Germany, Italy, Japan, the Netherlands, Spain, the United Kingdom, and the United States. See Barro and Sala-i-Martin (2004), Chapter 11.

[9]The exceptions are Cermeño (2002), who exploits the panel structure of the data, and Aroca, Bosch and Maloney (2003), who use an approach similar to the one used in this paper.

addresses one of the main criticisms of the functioning and capabilities literature;[10] (ii) as shown by several authors, improvements in health and increases in life expectancy have been some of the most important contributors to welfare over the past century;[11] and (iii) infant mortality and adult literacy are typically measured more accurately than per capita GDP, thus providing a better gauge with which to test the convergence hypothesis.

Second, this paper employs an alternative empirical methodology for studying the hypothesis of convergence to a common standard of living. This framework, known as the distribution dynamics approach, exploits both the time-series and cross-sectional dimensions of the data.

In particular, the methodology is based on tracking the evolution over time of the entire cross-section distributions across states through the estimation of transition matrices and kernel densities for relative per capita GDP, relative infant mortality and relative adult literacy, and analyzes changes over time in the external shape of these distributions as well as the intra distribution mobility of states.[12]

This approach allows one, under the Markov assumption, to obtain projection of the long-run distributions by computing the invariant or ergodic distributions implied by the estimated transition matrices. If there is convergence to a common standard of living, the distribution of each of the welfare indicators should converge to a mass point centered around the national average.

Thus, the convergence hypothesis is tested empirically by analyzing whether the invariant or ergodic distributions are increasingly concentrated around the national average. To the best of our knowledge, this approach has not been used before to analyze convergence to a common living standard.[13]

The distribution dynamics methodology was first proposed by Quah (1993) to analyze income convergence in a cross-section of countries. More recently, the approach has been employed to study regional income convergence by Bandyopadhyay (2002) for the case of states in India, by Lamo (2000) for the case of provinces in Spain, and by Magrini (1999) for the case of regions in the European Union.

In all of the cases above, the results obtained using the distribution dynamics approach challenge the notion that there has been a steady process of convergence among states and regions within each country.

This stands in contrast with the results of the earlier literature using regressions à la Barro, which suggested there was a relatively uniform rate of

[10]For an exposition of this approach, see Sen (2000).

[11]See, for example, Becker, *et al.* (2005), Murphy and Topel (2003), Nordhaus (2003) and Philipson and Soares (2002).

[12]By expressing a variable relative to its national average, it is possible to abstract from changes in the mean, which would be reflected in shifts in the distribution.

[13]Aroca, Bosch and Maloney (2003) use the same approach but highlight the spatial dimension of convergence in Mexico. For example, they analyze whether it is important for a state's growth performance whether its neighboring states are growing or not.

convergence (i.e. the so called iron law). Instead, more complex patterns of convergence emerge, including clustering into two or more groups.

This methodology has several advantages over the cross-country regression approach. First, at each point in time the procedure estimates the cross-section distributions non parametrically. Thus, it avoids the need to rely on any assumptions regarding the orthogonality of the covariates included in the regression equation and the disturbance term, or any assumption regarding the nature of long-run growth implicit in the averaging of growth rates over different time periods.

Second, by estimating the laws of motion of the entire cross-section distributions rather than estimating just some of their moments (e.g. their conditional mean or conditional variance), this approach permits the identification of richer patterns of convergence, including the possibilities of clustering into two groups (twin peaks), stratification or clustering into several groups (convergence clubs) or convergence to a common level.[14]

In order to estimate the transition matrices, this paper uses panel data from several sources. In the case of per capita GDP, it uses the decennial time series data on per capita real GDP by state constructed by Esquivel (1999) for the period 1940-1995, yearly time series data on real GDP by state from the National Institute of Statistics, Geography and Informatics[15] and on population by state from the National Population Council[16] for the period 1993-2002. In the case of the infant mortality rate and the adult literacy rate, it uses decennial census data from INEGI's General Population and Housing Censuses for the period 1940-2000.

The results show there is very low mobility —or, alternatively, high persistence— in the position states occupy in the distribution of per capita GDP, that mobility was highest between 1940 and 1950, and that it decreased steadily through 1980. In contrast, there is more mobility in the distributions of infant mortality and adult literacy, and mobility has increased throughout the period for the case of the infant mortality rate. Furthermore, there is no evidence in favor of the hypothesis of convergence to a common per capita GDP level or to a common infant mortality rate. In contrast, there is very clear evidence of convergence to a common adult literacy rate.

The rest of the paper is organized as follows. The next section presents a brief description of the Markov chain model on which the distribution dynamics approach is based. Section 3 describes the data sets used and discusses the results of the estimation of the transition matrices and the computation of the invariant distributions. Section 4 presents the main conclusions of the paper. Finally, an appendix presents a formal derivation of the frequency estimators used in the paper.

---

[14] See Durlauf and Quah (1999).

[15] Instituto Nacional de Estadística, Geografía e Informática, or INEGI for its acronym in Spanish.

[16] Consejo Nacional de Población, or CONAPO for its acronym in Spanish.

2.   MARKOV CHAINS

This section presents a brief review of Markov chains, a stochastic process widely used in economics to describe the mobility of agents or economic units across different states. The model employed in this paper is based on the discrete-time, finite-state Markov chain model first used by Quah (1993) to analyze income convergence across countries. The description of this model is based on Ljungqvist and Sargent (2004).

The relative per capita GDP of each economy in period t is represented by a random variable Y_t.[17]   The sequence of observations over time on relative per capita GDP $\{Y_t\}_{t \geq 1}$ is a stochastic process with a discrete time parameter. The first observation, $\overline{Y_1}$, is called the initial state of the process, and for $t = 2, 3,...$, the observation Y_t is called the state of the process at time t.

In each period t there are n mutually exclusive states, so an economy has to occupy one of these states. The probability model for relative per capita GDP is given by an initial $1 \times n$ probability vector $\boldsymbol{p_1}$, which describes the probability of the possible values of the initial state Y_1:

(1) $\boldsymbol{p}_{1i} = Pr(Y_1 = y_i), \quad \boldsymbol{p}_{1i} \geq 0 \ for \ i = 1,...,n, \ \sum_{t=1}^{n} \boldsymbol{p}_{1i} = 1$

and for each of the subsequent states Y_{t+1}, $t = 1, 2, 3,...$, every conditional probability of the form:

(2) $Pr(Y_{t+1} = y_{t+1} \mid Y_t = y_t, Y_{t-1} = y_{t-1},...,Y_1 = y_1)$

A Markov chain is a type of stochastic process such that for any time t, $t = 1, 2, 3,...$, and for any possible sequence of states $\{ y_1, y_2,...,y_t \}$:

(3) $Pr(Y_{t+1} = y_{t+1} \mid Y_t = y_t, Y_{t-1} = y_{t-1},...,Y_1 = y_1) = Pr(Y_{t+1} = y_{t+1} \mid Y_t = y_t)$

that is, the probability of all future states Y_{t+k}, $k \geq 1$, depends only on the current state Y_t and not on the previous states $\{Y_1, Y_2,..., Y_{t-1}\}$. The conditional probability $Pr(Y_{t+1} = y_{t+1} \mid Y_t = y_t)$ is called a transition probability. If transition probabilities have the same value for every time t, $t = 1, 2, 3,...$, then the Markov chain is said to have stationary transition probabilities:

(4) $Pr(Y_{t+1} = y_j \mid Y_t = y_i) = p_{ij}$

---

[17]Several clarifications are in order. First, in what follows relative per capita GDP is used as an example but may be replaced by the relative infant mortality rate or the relative adult literacy rate. Second, the term «economy» may refer to a state, province, region, or country. Third, the term «state» may refer to a state (or regime) of the stochastic process or to a geographic state, and proper care has been taken to distinguish between them.

where $p_{ij} \geq 0$ is the probability that an economy will be in state j next period given that it is in state i this period. In this case, we can represent the probabilities of moving from one value of the state to another in one period using an $n \times n$ one-step transition matrix P:

$$(5) \quad P_{n \times n} = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ p_{1n} & \cdots & p_{nn} \end{pmatrix}$$

Since an economy has to occupy a state next period, regardless of the state it occupies this period, each row of the matrix must satisfy:

$$(6) \quad \sum_{j=1}^{n} p_{ij} = 1, \text{ for } i = 1, \dots, n$$

If a Markov chain has stationary transition probabilities, given a transition matrix P for a single step, we can compute the probabilities of moving from any value of the state to any other value of the state in two periods as P^2, since:

$$(7) \quad Pr(Y_{t+2} = y_j \mid Y_t = y_i) = \sum_{h=1}^{n} Pr(Y_{t+1} = y_j \mid Y_t = y_h) Pr(Y_{t+1} = y_h \mid Y_t = y_i)$$

$$= \sum_{h=1}^{n} p_{ih}\, p_{hj} = p_{ij}^2$$

Similarly, we can compute the probabilities of moving from any value of the state to any other value of the state in k periods as P^k:

$$(8) \quad Pr(Y_{t+k} = y_j \mid Y_t = y_i) = p_{ij}^k$$

The unconditional probability distribution of Y_t is given by:

$$(9) \quad \begin{aligned} \boldsymbol{p}_2 &= Pr(Y_2) = \boldsymbol{p}_1 P \\ \boldsymbol{p}_3 &= Pr(Y_3) = \boldsymbol{p}_1 P^2 = \boldsymbol{p}_2 P \\ &\cdot \\ &\cdot \\ &\cdot \\ \boldsymbol{p}_k &= Pr(Y_k) = \boldsymbol{p}_1 P^k = \boldsymbol{p}_{k-1} P \end{aligned}$$

From the above equations we can see that the unconditional probability distribution evolves according to:

$$\boldsymbol{p}_{t+1} = \boldsymbol{p}_t P$$

A distribution is called invariant or ergodic if:

$$\boldsymbol{p}_{t+1} = \boldsymbol{p}_t = \boldsymbol{p}^* \text{ forall } t$$

that is, the unconditional distribution remains constant over time. Thus, an invariant or ergodic distribution must satisfy:

$$p^* = p^* P$$

$$p^*(I - P) = 0$$

Transposing this last equation yields:

(10) $(I - P')p^{*\prime} = 0$

which determines p^* as an eigenvector (normalized to satisfy $\sum_{i=1}^{n} p_i = 1$) associated with a unit eigenvalue of P'. The fact that P is a stochastic matrix guarantees that it has at least one unit eigenvalue, and that there is some p^* that satisfies $(I - P')p^{*\prime} = 0$. Depending on P, an invariant distribution may or may not be unique. In particular, if every entry of the matrix P is strictly positive, then there exists a unique invariant distribution called the stationary distribution.

3.   DATA SOURCES, ESTIMATION, AND COMPUTATION

This section describes the data sources and presents the results from the estimation of the empirical transition matrices $\hat{P}$, the computation of their implied invariant or ergodic distributions p^*, and the estimation of the bivariate kernel densities (stochastic kernels).

3.1.   Data

The data for state GDP used are drawn from two sources. The first is the series constructed by Esquivel (1999). This data set constitutes the longest series available for state-level GDP in Mexico. It is a panel consisting of observations on real per capita GDP for all 31 states and the Federal District every 10 years for the period 1940-1990 and for 1995.

The second is the yearly data from INEGI. It is also a panel consisting of observations on real GDP for all 31 states and the Federal District every year for the period 1993-2002. In order to convert them into per capital GDP, each state's GDP is divided by the state's population provided by the yearly projections by CONAPO. In this paper the two data sets are combined to produce a decennial series for the period 1940-2000.

Most of the changes over time in a state's per capita GDP are the result of the relative performance of each state. Nevertheless, given the sharp changes observed in the relative per capita GDP in the cases of the states of Campeche and Tabasco between 1960, 1970 and 1980, a clarification is in order. By analyzing more disaggregated state-level GDP data (i.e., by sectors of economic activity), it can be easily verified that most of the changes are due to differences in the state income and product accounting methodology.

In particular, the way in which value added from crude oil and natural gas extraction was attributed to each state changed over time. Most of the oil in Mexico is located off the coast of these two states, but since it is owned and exploited by the federal state-oil monopoly (Pemex) it should not be considered as part of their income. In order to circumvent this problem, all estimations and projections were repeated with and without these states and no major differences were found.

In the case of the infant mortality rate and the adult literacy rate, the data come from INEGI's decennial General Population and Housing Censuses for the period 1940-2000.[18]   The infant mortality rate is defined as the number of deaths occurred among children less than one year of age for each 1,000 live births, and the adult literacy rate is defined as the percentage of the population over a given age (typically 15) who can read and write a short message.[19]

3.2.   Estimation and Computation

We are interested in the evolution over time of relative per capita GDP, the relative infant mortality rate, and the relative adult literacy rate across states, and on their invariant or ergodic distributions as a way of characterizing their long-run distributions.

If there is convergence in these welfare indicators, we should observe their distributions converging to a mass point centered around the national average. If there is no convergence, however, this approach allows us to identify richer patterns of convergence, including clustering into two groups (twin peaks) or stratification or clustering into several groups (convergence clubs), through analyzing the shape of the invariant distribution.

The assumption made for computing the invariant or ergodic distributions is that relative per capita GDP, the relative infant mortality rate, and the relative adult literacy rate follow a Markov process with stationary transition probabilities. This implies that the value of each of these welfare indicators at any given time t depends only on their values in period $t-1$, respectively, and not on any of the previously observed values.

In reality, however, relative per capita GDP, the relative infant mortality rate, and the relative adult literacy rate at times $t-k, k \geq 2$, may provide useful information in determining their values at time t. So whether the Markov assumption holds is an empirical matter which depends on the length of the interval t; the assumption may be unreasonable for short periods of time such as one year, but reasonable for longer periods such as a decade.

Following Quah (1993), we consider per capita GDP, the infant mortality rate, and the adult literacy rate in each (geographic)\ state relative to their national

---

[18]See INEGI (2000b) and INEGI (2001).

[19]The reason for the choice of these two welfare indicators instead of more preferable measures such as secondary school enrollment and life expectancy at birth is their availability and that their measurement is consistent throughout the different censuses.

averages as the basic data, and define the states of the process as intervals. The national average for each of these three welfare indicators is a weighted average of each state's value, where each state's weigh is the state's population share.

We first discretize the set of possible values of relative per capita GDP into the following five intervals: (0,0.5), [0.5,0.75], [0.75,1.25], (1.25,1.5] and (1.5, ∞). In the case of the adult literacy rate, we use the following five intervals: (0,0.925), [0.925,0.975), [0.975,1.025], (1.025,1.075], (1.075, ∞). As for the infant mortality, the intervals used are: (0,0.85), [0.85,0.95), [0.95,1.05], (1.05,1.15], (1.15, ∞).

Notice that these intervals are different for the three welfare measures, and in no case are they equally-sized. In all cases they were chosen, if somewhat arbitrarily, so that two conditions were satisfied: (i) there is an odd number of intervals and the middle interval is centered around 1; and (ii) for every estimated matrix there is always at least one observation in each state of the process.

The procedure by which we estimate the transition matrices is non parametric. In particular, each entry of the 5 x 5 transition matrix is estimated as the empirical frequency in the sample; i.e., the number of (geographic) states that had per capita GDP, adult literacy or infant mortality in the same given interval in two periods, divided by the number of (geographic) states that began with per capita GDP, adult literacy or infant mortality, respectively, in the given interval.

This empirical frequency estimator in fact corresponds to the maximum likelihood estimator, as shown in the appendix. Thus, the estimator of the stationary transition probability p_{ij} is given by:

$$(11) \qquad \hat{p}_{ij} = \frac{\sum_{h=1}^{32} d_{hij}}{\sum_{h=1}^{32} d_{hi}}$$

where:

$$d_{hij} = \begin{cases} 1 \text{ if } y_{h,t+1} \in I_j \text{ and } y_{h,t} \in I_i \\ 0 \text{ otherwise} \end{cases}$$

$$d_{hi} = \begin{cases} 1 \text{ if } y_{h,t} \in I_i \\ 0 \text{ otherwise} \end{cases}$$

and the summation is over all (geographic) states h, $h = 1$,..., 32.

Notice that the values of the states of the process are not defined as relative per capita GDP, the relative infant mortality rate, and the relative adult literacy rate, which are continuous variables, but as indicator variables which classify (geographic) states according to the different intervals I in which the value of their indicators is found each period.

Since we are also interested in the long-run GDP distribution to analyze the issue of convergence to a common standard of living, we compute the invariant or

ergodic distributions implied by the estimated matrices using the method described in the previous section.

The invariant or ergodic distributions are computed for the one-step transition matrices covering the longest interval (1940-2000). The results of the estimation are shown in Tables 1 through 3, where the numbers on the left of $\hat{P}$ represent the number of (geographic) states in each interval and the vector below the last matrix is the computed invariant or ergodic distribution.

TABLE 1
ESTIMATED ONE-STEP TRANSITION MATRICES AND INVARIANT
DISTRIBUTION FOR REAL PER CAPITA GDP RELATIVE TO THE
NATIONAL AVERAGE: 1940-2000
Intervals: (0,0.5), [0.5,0.75), [0.75,1.25], (1.25,1.5), (1.5, ∞)

1940-1950

4	0.750	0.250	0.000	0.000	0.000
9	0.111	0.556	0.333	0.000	0.000
8	0.000	0.125	0.875	0.000	0.000
4	0.000	0.000	0.500	0.250	0.250
7	0.000	0.286	0.143	0.000	0.571

1950-1960

4	0.750	0.250	0.000	0.000	0.000
9	0.222	0.444	0.333	0.000	0.000
13	0.000	0.308	0.692	0.000	0.000
1	0.000	0.000	0.000	0.000	1.000
5	0.000	0.000	0.200	0.200	0.600

1960-1970

5	0.800	0.200	0.000	0.000	0.000
9	0.222	0.556	0.222	0.000	0.000
13	0.000	0.154	0.692	0.154	0.000
1	0.000	0.000	1.000	0.000	0.000
4	0.000	0.000	0.000	0.000	1.000

1970-1980

6	0.667	0.167	0.167	0.000	0.000
8	0.000	0.875	0.000	0.000	0.125
12	0.000	0.167	0.750	0.083	0.000
2	0.000	0.000	1.000	0.000	0.000
4	0.000	0.000	0.000	0.500	0.500

1980-1990

4	0.500	0.500	0.000	0.000	0.000
10	0.000	0.700	0.200	0.000	0.100
12	0.083	0.000	0.750	0.167	0.000
3	0.000	0.000	0.000	1.000	0.000
3	0.000	0.000	0.333	0.000	0.667

1990-1995

3	0.667	0.333	0.000	0.000	0.000
9	0.000	1.000	0.000	0.000	0.000
12	0.000	0.167	0.750	0.083	0.000
5	0.000	0.000	0.000	0.800	0.200
3	0.000	0.000	0.000	0.000	1.000

1990-2000

3	0.667	0.333	0.000	0.000	0.000
9	0.000	0.778	0.222	0.000	0.000
12	0.000	0.167	0.000	0.000	0.000
5	0.000	0.000	0.200	0.800	0.000
3	0.000	0.000	0.000	0.000	1.000

1940-2000

4	0.500	0.500	0.000	0.000	0.000
9	0.000	0.667	0.222	0.000	0.111
8	0.000	0.250	0.750	0.000	0.000
4	0.000	0.000	0.250	0.500	0.250
7	0.000	0.000	0.571	0.286	0.143

Invariant distribution: 1940-1990

0.000	0.388	0.517	0.034	0.060

TABLE 2

ESTIMATED ONE-STEP TRANSITION MATRICES AND INVARIANT
DISTRIBUTION FOR THE INFANT MORTALITY RATE TO THE
NATIONAL AVERAGE: 1940-2000

Intervals: (0,0.85), [0.85,0.95], [0.95,1.05], (1.05,1.15] (1.15, ∞)

1940-1950

5	0.400	0.600	0.000	0.000	0.000
13	0.231	0.692	0.077	0.000	0.000
7	0.000	0.286	0.714	0.000	0.000
6	0.000	0.000	0.333	0.333	0.333
1	0.000	0.000	0.000	0.000	1.000

1950-1960

5	0.600	0.400	0.000	0.000	0.000
14	0.071	0.571	0.214	0.143	0.000
8	0.000	0.000	0.750	0.250	0.000
2	0.000	0.000	0.000	1.000	0.000
3	0.000	0.000	0.000	0.333	0.667

1960-1970

4	0.750	0.250	0.000	0.000	0.000
10	0.200	0.700	0.100	0.000	0.000
9	0.000	0.333	0.444	0.222	0.000
7	0.000	0.143	0.429	0.286	0.143
2	0.000	0.000	0.000	0.000	1.000

1970-1980

5	0.600	0.000	0.400	0.000	0.000
12	0.333	0.417	0.167	0.083	0.000
8	0.000	0.250	0.500	0.125	0.125
4	0.000	0.000	0.250	0.500	0.250
3	0.000	0.000	0.000	0.000	1.000

1980-1990

7	0.714	0.286	0.000	0.000	0.000
7	0.571	0.143	0.286	0.000	0.000
9	0.000	0.667	0.222	0.111	0.000
4	0.000	0.500	0.500	0.000	0.000
5	0.000	0.000	0.000	0.200	0.800

1990-1995

9	0.778	0.222	0.000	0.000	0.000
11	0.091	0.182	0.545	0.182	0.000
6	0.000	0.167	0.667	0.167	0.000
2	0.000	0.000	0.500	0.500	0.000
4	0.000	0.000	0.000	0.000	1.000

1990-2000

9	0.667	0.333	0.000	0.000	0.000
11	0.000	0.364	0.545	0.091	0.000
6	0.167	0.000	0.333	0.333	0.167
2	0.000	0.000	0.500	0.500	0.000
4	0.000	0.000	0.000	0.000	1.000

1940-2000

5	0.400	0.000	0.400	0.200	0.000
13	0.308	0.308	0.308	0.000	0.077
7	0.143	0.429	0.286	0.143	0.000
6	0.000	0.000	0.167	0.167	0.667
1	0.000	0.000	0.000	1.000	0.000

Invariant distribution 1940-2000

0.106 0.118 0.191 0.346 0.239

TABLE 3
ESTIMATED ONE-STEP TRANSITION MATRICES AND INVARIANT
DISTRIBUTION FOR THE ADULT LITERACY RATE RELATIVE TO THE
NATIONAL AVERAGE: 1940-2000
Intervals: (0,0.925), [0.925,0.975], [0.975,1.025], (1.025,1.075], (1.075, ∞)

1940-1950

12	0.917	0.000	0.083	0.000	0.000
1	0.000	1.000	0.000	0.000	0.000
1	0.000	0.000	0.000	1.000	0.000
1	0.000	0.000	0.000	1.000	0.000
17	0.000	0.000	0.000	0.059	0.941

1950-1960

11	1.000	0.000	0.000	0.000	0.000
1	0.000	1.000	0.000	0.000	0.000
1	0.000	0.000	1.000	0.000	0.000
3	0.000	0.333	0.000	0.667	0.000
16	0.000	0.000	0.000	0.250	0.750

1960-1970

11	0.727	0.182	0.091	0.000	0.000
2	0.000	0.000	1.000	0.000	0.000
1	0.000	0.000	1.000	0.000	0.000
6	0.000	0.167	0.167	0.667	0.000
12	0.000	0.000	0.083	0.083	0.833

1970-1980

8	1.000	0.000	0.000	0.000	0.000
3	0.000	0.667	0.333	0.000	0.000
6	0.000	0.000	0.833	0.167	0.000
5	0.000	0.000	0.200	0.800	0.000
10	0.000	0.000	0.000	0.000	1.000

1980-1990

8	0.500	0.500	0.000	0.000	0.000
2	0.000	0.500	0.500	0.000	0.000
7	0.000	0.286	0.714	0.000	0.000
5	0.000	0.000	0.000	1.000	0.000
10	0.000	0.000	0.000	0.300	0.700

1990-1995

4	0.750	0.250	0.000	0.000	0.000
7	0.000	0.857	0.143	0.000	0.000
6	0.000	0.167	0.833	0.000	0.000
8	0.000	0.000	0.375	0.625	0.000
7	0.000	0.000	0.000	0.714	0.286

1990-2000

4	0.750	0.250	0.000	0.000	0.000
7	0.000	0.571	0.429	0.000	0.000
6	0.000	0.000	1.000	0.000	0.000
8	0.000	0.000	0.250	0.750	0.000
7	0.000	0.000	0.000	0.857	0.143

1940-2000

12	0.250	0.333	0.333	0.083	0.000
1	0.000	0.000	1.000	0.000	0.000
1	0.000	0.000	1.000	0.000	0.000
1	0.000	0.000	1.000	0.000	0.000
17	0.000	0.059	0.235	0.647	0.059

Invariant distribution 1940-2000

0	0	1	0	0

Several results emerge from the estimation of the transition matrices. First, there is low mobility (alternatively, high persistence) among states in the position they occupy in the distribution of relative per capita GDP, even in the long run (1940-2000), as compared to the mobility in the distributions of relative adult literacy and relative infant mortality. This fact manifests itself in the higher value of the diagonal terms in the estimated transition matrices and in the values of the mobility indexed calculated from them.

In particular, for each estimated transition matrix we computed the following four mobility indexed: Shorrocks's harmonic mean index, Geweke, Marshall and Zarkin's Eigenvalue index, Sommers and Conlisk's Second Largest Eigenvalue index, and the Implied Auto-Regressive Coefficient. All of these indexes summarize

in a scalar the mobility within the distribution of each welfare indicator. The results of these indexes are shown in Figures 1, 2, and 3, which show the values of the mobility indexes for each of the estimated one-step transition matrices.[20]

FIGURE 1
MOBILITY INDEXES FOR REAL PER CAPITA GDP
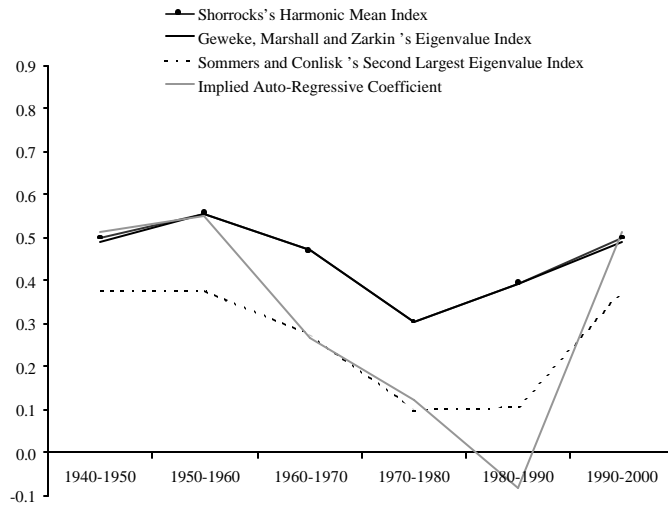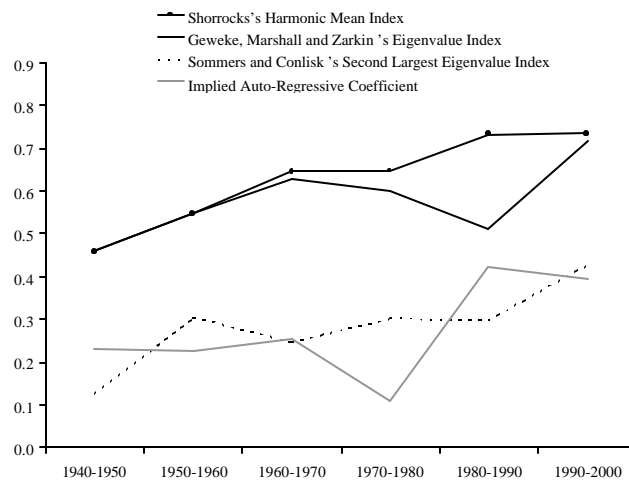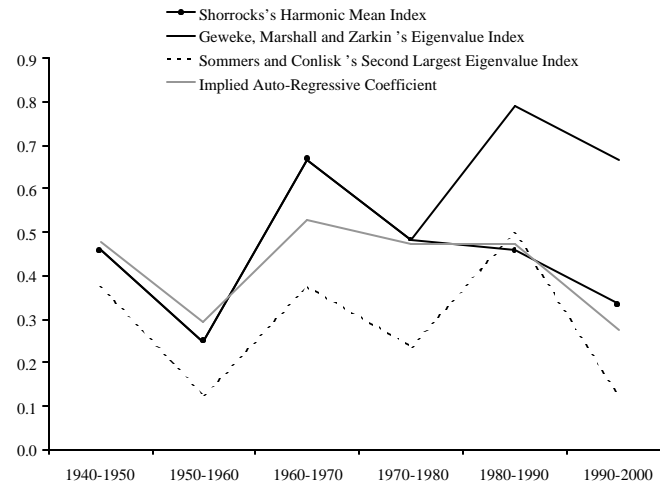RELATIVE TO THE NATIONAL AVERAGE 1940-200

FIGURE 2
MOBILITY INDEXES FOR THE INFANT MORTALITY RATE
RELATIVE TO THE NATIONAL AVERAGE: 1940-200

[20]See Geweke, *et al.* (1986) for more details on the construction of these mobility indexes.

FIGURE 3
MOBILITY INDEXES FOR THE ADULT LITERACY RATE
RELATIVE TO THE NATIONAL AVERAGE: 1940-200

Second, the estimated matrices display more persistence in relative per capita GDP the shorter the interval, as expected. This can be seen by comparing the matrix for 1940-2000 with any of the matrices for any of the consecutive decades.

Third, there is generally a higher probability of falling behind than moving ahead the national average, as shown by the diagonal and off-diagonal terms of the lowest and highest fractiles.

Fourth, as for relative per capita GDP and the infant mortality rate in the long run, their implied invariant or ergodic distributions show no evidence in favor of the convergence hypothesis. While most of the probability mass (around 80%) concentrates around the national average income, there is still a significant probability of being below half the national average and of being above one and a half times the national average.

There is no evidence, however, of clustering into two or more income groups (the so called twin peaks and convergence clubs hypotheses), which would be reflected in the relative income distribution converging to a bimodal or multimodal invariant distribution. For the case of the adult literacy rate, there is strong evidence in favor of the absolute convergence hypothesis. This can be seen in the computed invariant distribution, which has all the mass concentrated in the middle interval containing the national average.

These results are only meant to be suggestive, since the choice of intervals (states) for each variable is arbitrary and different sets of discretizations may lead to different invariant distributions.[21]

[21]See Bulli (2001).

In order to address this issue, we repeat this analysis using the continuous-state version of the discrete-state, transition matrix approach is needed to assure that these findings are robust to the choice of intervals. This approach is based on estimating a bivariate kernel density, or stochastic kernel, for each pair of decades. The kernel used in all cases is the Gaussian kernel.

The results from these estimations can be seen in Figures 4, 5 and 6. These figures depict the kernel density estimates for each pair of consecutive decades, for the five year period 1990-1995, as well as for the longest period available (1940-2000). The graphs on the left-hand side depict the bivariate probability density functions, while the graphs on the right-hand side depict a series of level sets for these density functions.

The X and Y axes represent the states' per capita GDP, infant mortality and adult literacy, respectively, each expressed relative to their national averages. It is important to notice that in all cases the national averages are depicted in each axis by the horizontal and vertical lines at one.[22] The area above the (X, Y) plane represents the probability density function.

These figures convey a large amount of information in a very condensed way. First, the graphs seen as a sequence can be used to determine whether convergence to a common level of each variable has occurred over the period analyzed. To the extent that the sequence of graphs display an increasing concentration over time of the mass of the density function around the national average, then this can be interpreted as evidence in favor of the absolute convergence hypothesis.[23]

Second, the last graph of each sequence represents the kernel density estimate for the longest period available (1940-2000), and can also be used to determine whether convergence to a common level of each variable has occurred. In this case convergence would manifest itself if all the states' relative variables are clustered around their national average in 2000, independent of the position they had in 1940. Thus, the relative variances of the data on the X and Y axes also help in determining the extent of convergence.

Third, the graphs in the right-hand side panels in Figures 4, 5 and 6 can also be used to determine the extent of mobility alluded before . To the extent that observations cluster around the 45-degree line, there is little or no mobility since their relative position is the same in both decades.

The analysis of the kernel density estimates confirms the results of the estimated transition matrices and the associated invariant distribution. In particular, the sequences of estimated densities for per capita GDP and the infant mortality rate show no evidence of convergence, while the one for the adult literacy rate shows very clear evidence of convergence.

[22]Recall that the national average is one by definition since each of the three welfare indicators is expressed relative to its national average.

[23]The mass should be centered around the point (1,1), the average for both periods.

FIGURE 4
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF REAL PER CAPITA GDP RELATIVE TO THE
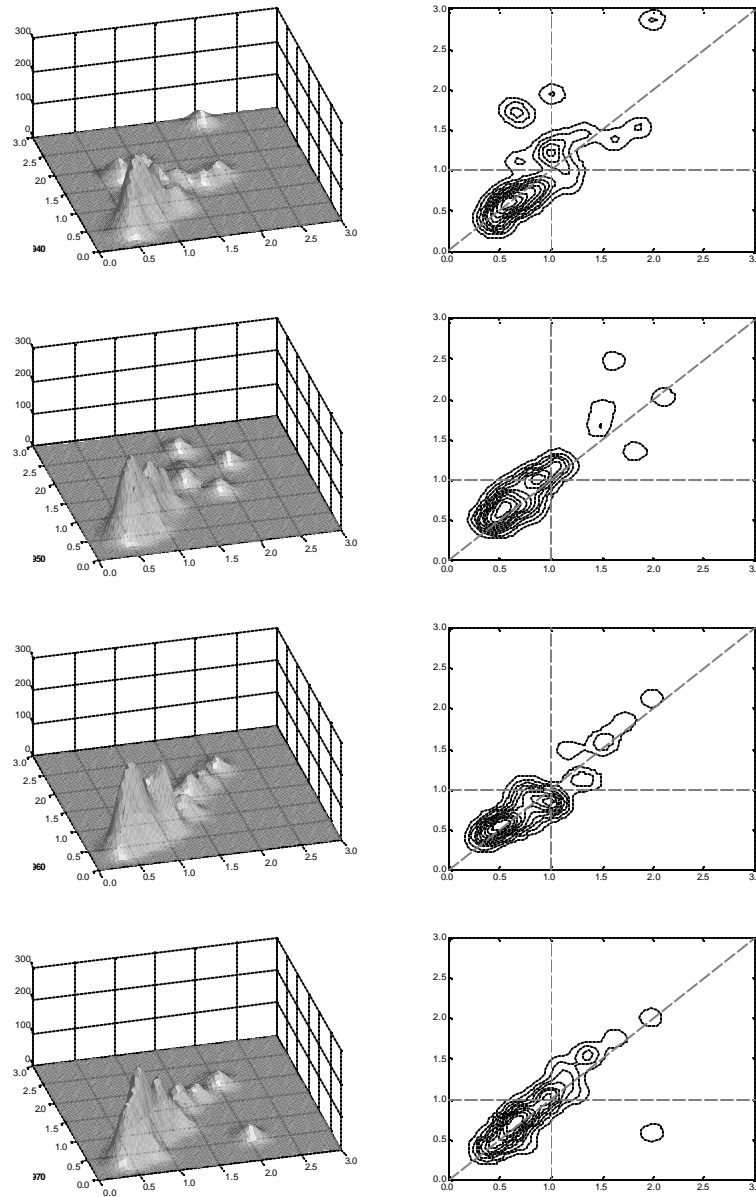NATIONAL AVERAGE: 1940-2000

FIGURE 4 (CONTINUED)
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF REAL PER CAPITA GDP RELATIVE TO THE
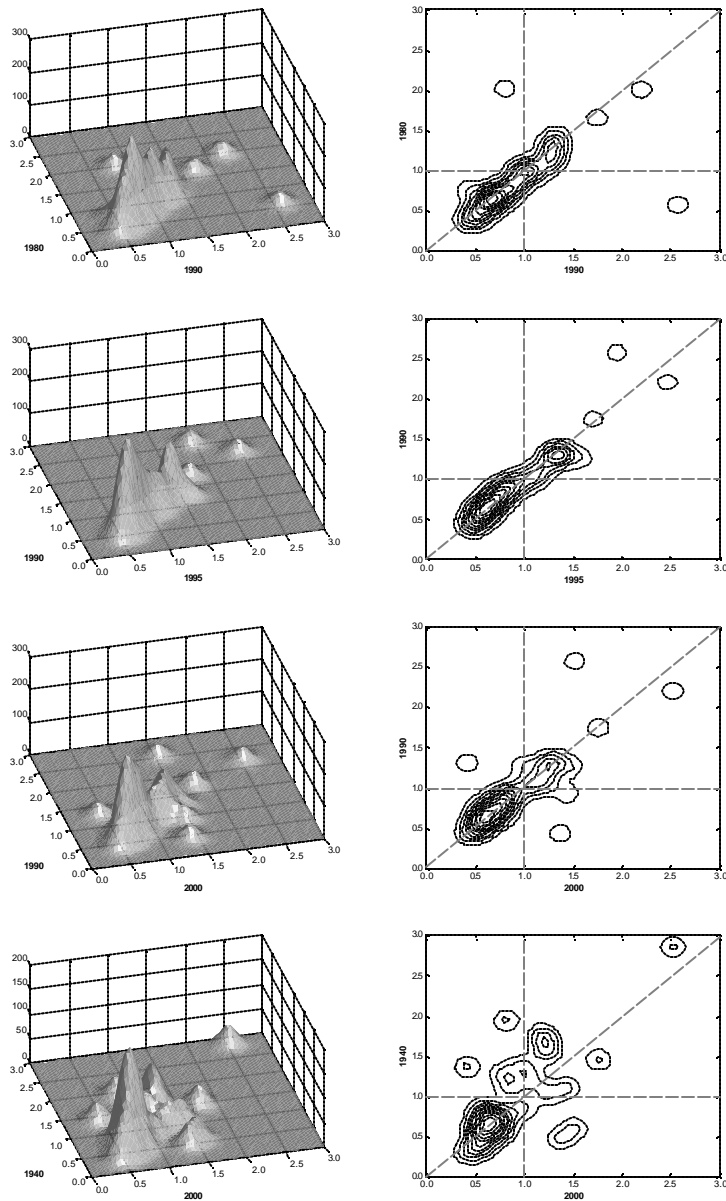NATIONAL AVERAGE: 1940-2000

FIGURE 5
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF THE INFANT MORTALITY RATE RELATIVE TO THE
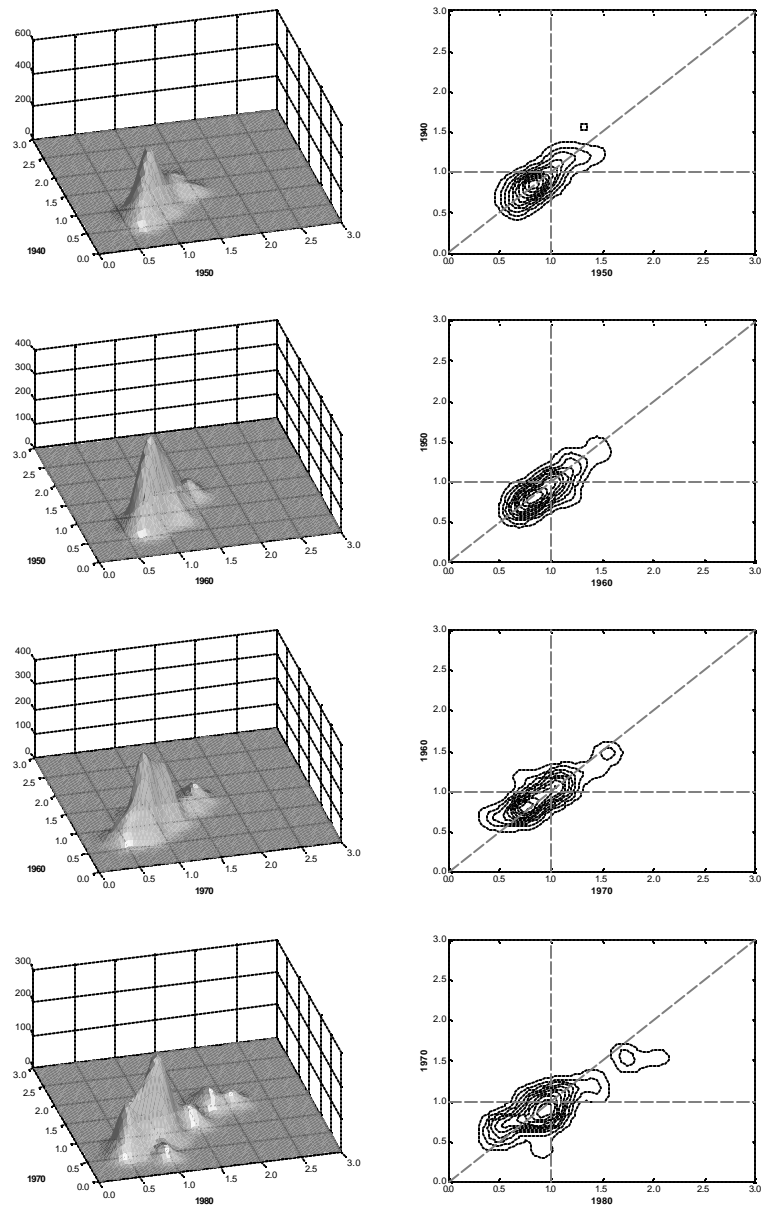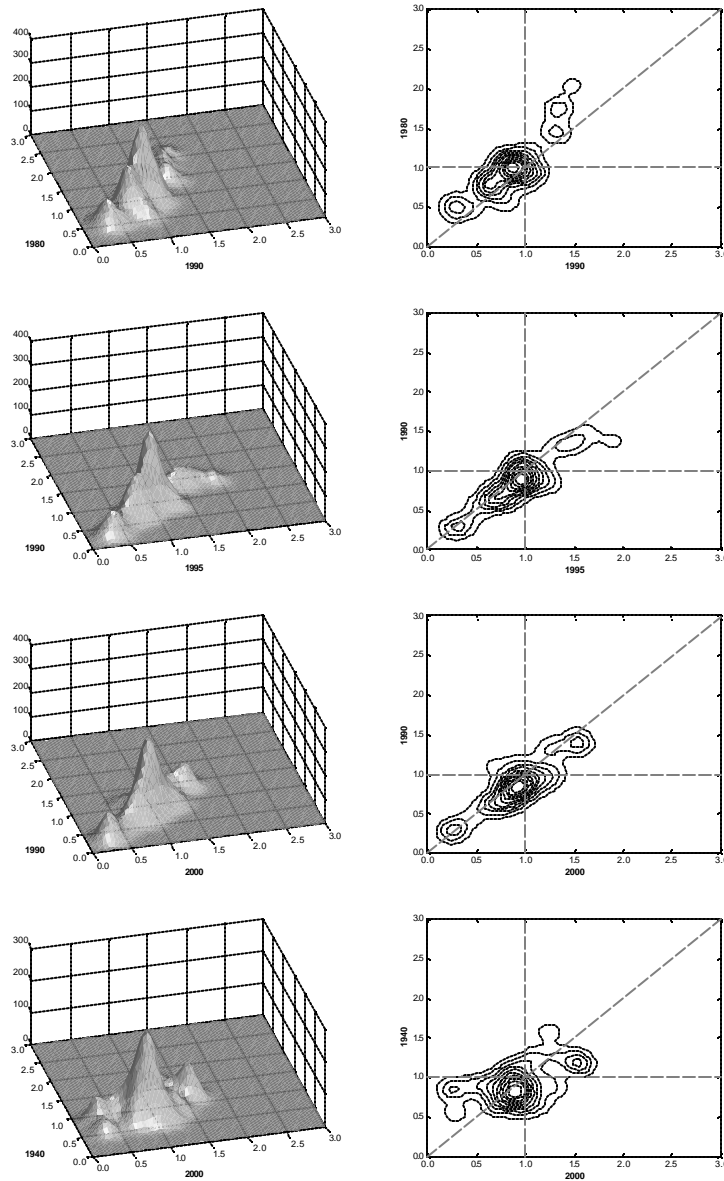NATIONAL AVERAGE: 1940-2000

FIGURE 5 (CONTINUED)
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF THE INFANT MORTALITY RATE RELATIVE TO THE
NATIONAL AVERAGE: 1940-2000

FIGURE 6
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF THE ADULT LITERACY RATE RELATIVE TO THE
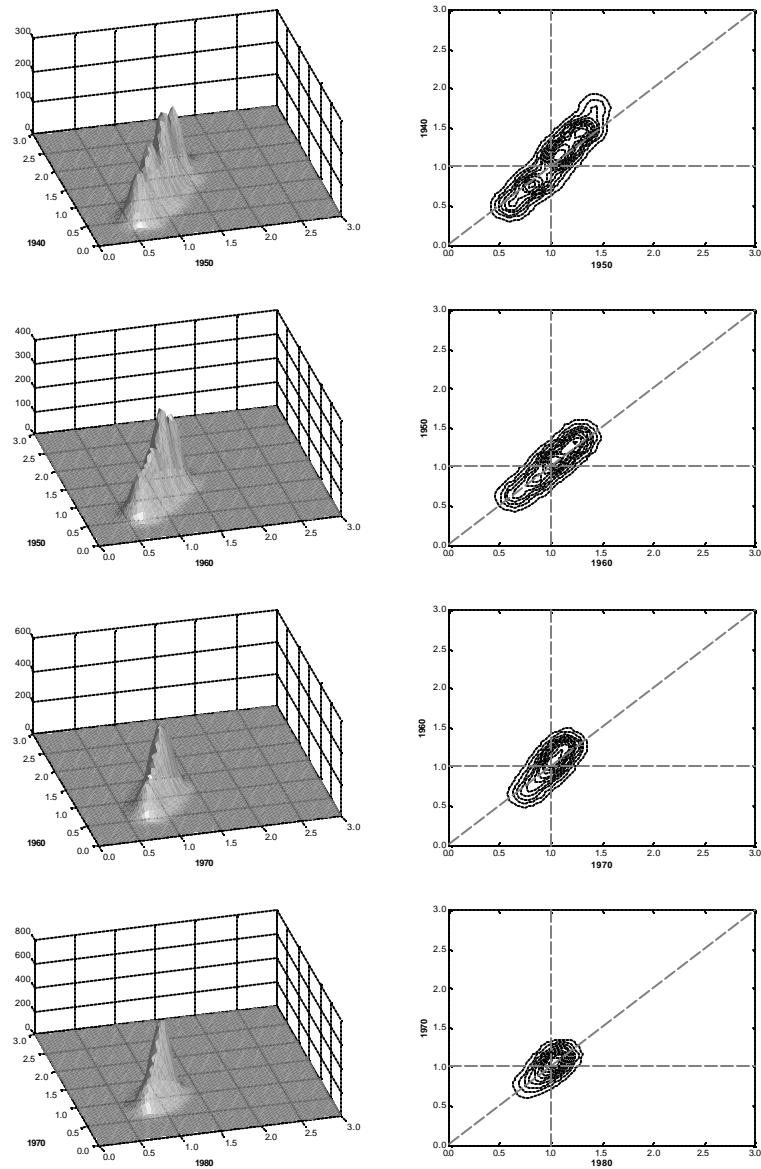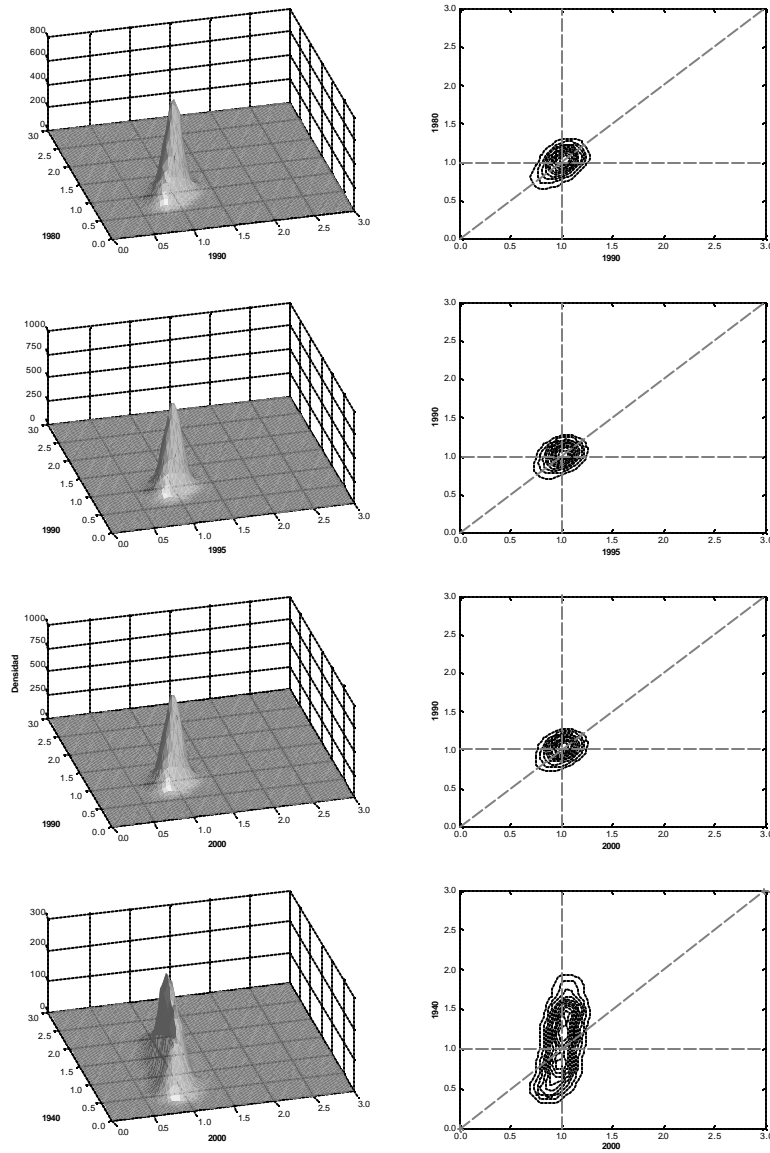NATIONAL AVERAGE: 1940-2000

FIGURE 6 (CONTINUED)
BIVARIATE KERNEL DENSITY ESTIMATES (STOCHASTIC KERNELS)
OF THE ADULT LITERACY RATE RELATIVE TO THE
NATIONAL AVERAGE: 1940-2000

These densities also confirm that while there has not been a process of convergence, there has not been a process of stratification either, which would be reflected in the clustering of states into two or more peaks.

As mentioned before, this confirmation is very important given that the discretization of the state-space for each variable in arbitrary intervals affects the implied invariant distributions.

4.   CONCLUSIONS

The results of the estimation provide no evidence in support of the hypothesis of convergence among Mexican states to a common per capita GDP level or to a common infant mortality rate. In the context of a Markov chain model, convergence would manifest itself in an invariant or ergodic distribution with a mass point centered around the national average.

In the case of per capita GDP, we observe that most of the mass of the invariant distribution concentrates around the national average income. In particular, the probability that a state will be between 0.75 (0.5) and 1.25 (1.5) times the national average is around 44% (80%). Nevertheless, there is a probability of about 13% that a state will be at less than half the national average and a probability of about 9% that a state will be at more than one and a half times the national average.

This result, together with the low mobility displayed by the estimated transition matrices, implies there is no convergence among these states to a common per capita GDP level, so that the poorest states will remain poor while the richest states will remain rich. In contrast with Quah (1993), who found that the world income distribution has tended towards clustering into two income groups (i.e., a bimodal distribution), we find that the invariant distribution is symmetric.

In the case of the infant mortality rate, most of the mass of the invariant distribution concentrates around the national average. In particular, the probability that a state will be between 0.9 (0.7) and 1.10 (1.3) times the national average is also around 46% (80%). Nevertheless, there is a probability of approximately 3% that a state will be at less than 0.7 the national average and a probability of approximately 14% that a state will be at more than 1.3 times the national average. We also find, as with per capita GDP, that the invariant distribution is fairly symmetric.

As for the mobility of states within the infant mortality rate distribution, we find there is more mobility than in the case of per capita GDP, and that this mobility has been increasing over time. Thus, although the external shape of the distribution has been fairly constant over time, it is more likely that states will change their position inside the distribution. Finally, the invariant distribution of the adult literacy rate is the only one consistent with the absolute convergence hypothesis, since all the mass is concentrated around the center interval containing the national average.

It is important to highlight that the results obtained from the estimated transition matrices and their implied invariant distributions are in all cases confirmed

by the kernel density estimates, which is important given the arbitrary nature of the state-space for the estimation of the transition matrices and their effect on the invariant distributions.

5.   REFERENCES

Aroca, P.; M. Bosch and W. Maloney (2003), "Is NAFTA Polarizing Mexico or ¿El Sur También Existe? Spatial Dimensions of Mexico's Post-Liberalization Growth," mimeo, The World Bank, January.

Bandyopadhyay, S. (2002), "Twin Peaks: Distribution Dynamics of Economic Growth across Indian States", mimeo, London School of Economics, January.

Barro, R. J. (1997), *Determinants of Economic Growth: A Cross-Country Empirical Study*, MIT Press.

Barro, R. J. (1991), "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics*, 106 (2): 407-443.

Barro, R. J. and X. Sala-i-Martin (2004), *Economic Growth*, Second Edition, MIT Press: Cambridge, MA.

Barro, R. J. and X. Sala-i-Martin (1992), "Convergence", *Journal of Political Economy*, 100 (2): 223-251.

Barro, R. J. and X. Sala-i-Martin (1991), "Convergence across States and Regions", *Brookings Papers on Economic Activity*, 1:107-182.

Becker, G. S.; T. J. Philipson and R. R. Soares (2003), "The Quantity and Quality of Life and the Evolution of World Inequality", NBER Working Paper No. 9765, Cambridge, MA.

Bulli, S. (2001), "Distribution Dynamics and Cross-Country Convergence: A New Approach", *Scottish Journal of Political Economy*, 48 (2): 226-243.

Caraza, M.I. (1993), "Convergencia del ingreso en la República Mexicana," unpublished B.A. thesis, Instituto Tecnológico Autónomo de México, México City, México.

Cermeño, R. (2001), "Decrecimiento y Convergencia en los Estados Mexicanos: Un Análisis de Panel,'' El *Trimestre Económico*, 68 (4): 603-629.

Chiquiar, D. (2005), "Why Mexico's Regional Income Convergence Broke Down?, unpublished manuscript , University of California San Diego, forthcoming, *Journal of Development Economics*.

Durlauf, S. N. and D. Quah (1999), "The New Empirics of Economic Growth", in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*, 1a, Ch. 4, North-Holland.

Esquivel, G. (1999), "Convergencia Regional en México, 1940-1995,'' Documento de Trabajo 9, Centro de Estudios Económicos, El Colegio de México.

Geweke, J.; R. C. Marshall and G. A. Zarkin (1986), "Mobility Indices in Continuous Time Markov Chains", *Econometrica*, 54 (6): 1407-1423.

Heston, A.; R. Summers and B. Aten (2002), *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP).

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2003), *Sistema de Cuentas Nacionales de México (SCNM). Producto Interno Bruto por Entidad Federativa. 1997-2002*, México.

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2001), *XII Censo General de Población y Vivienda, 2000*, México.

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2000a), *Sistema de Cuentas Nacionales de México (SCNM). Producto Interno Bruto por Entidad Federativa. 1993-1996*, México.

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2000b), *Estadísticas Históricas de México. Tomo I*, México.

Jones, Ch. I. (1997), "On the Evolution of the World Income Distribution''. *Journal of Economic Perspectives*, 11 (3): 19-36.

Juan-Ramón, V. H. and L. A. Rivera-Batiz (1996), "Regional Growth in Mexico, 1970-93,'' IMF Working Paper, WP/96/92, International Monetary Fund.

Lamo, A. (2000), "On Convergence Empirics: Some Evidence from Spanish Regions,'' *Investigaciones Económicas*, 24 (3): 681-707.

Ljungqvist, L. and T. J. Sargent (2004), *Recursive Macroeconomic Theory*, 2nd edition, Cambridge, MA: The MIT Press.

Magrini, S. (1999), "The Evolution of Income Disparities Among Regions of the European Union,'' *Regional Science and Urban Economics*, 29 (2): 257-281.

Messmacher L., M. (2000), "Desigualdad regional en México. El Efecto del TLCAN y Otras Reformas Estructurales," Documento de Investigación, Nr. 2000-4, Dirección General de Investigación Económica, Banco de México, December.

Murphy, K. M. and R. H. Topel (2003), "The Economic Value of Medical Research," in *Measuring the Gains from Medical Research: An Economic Approach*, Kevin M. Murphy and Robert H. Topel, eds., The University of Chicago Press, Chicago, IL.

Navarrete, J. (1995), "Convergencia: Un Estudio para los Estados de la República Mexicana," Documento de Trabajo del CIDE, División de Economía, Nr. 42, México.

Nordhaus, W. D. (2003), "The Health of Nations: The Contribution of Improved Health to Living Standards," in *Measuring the Gains from Medical Research: An Economic Approach*, Kevin M. Murphy and Robert H. Topel, eds., The University of Chicago Press, Chicago, IL.

Parente, S. L. and E. C. Prescott (1993), "Changes in the Wealth of Nations,'' *Federal Reserve Bank of Minneapolis Quarterly Review*, 17 (2) unnumbered.

Philipson, T. J, and R. R. Soares (2002), "World Inequality and the Rise in Longevity," in *Annual World Bank Conference on Development Economics 2001/2002*, Boris Pleskovic and Nicholas Stern, eds., The World Bank and Oxford University Press, Washington, DC, 245-259.

Pritchett, L. (1997), "Divergence, Big Time''. *Journal of Economic Perspectives*, 11 (3): 3-17.

Quah, D. (1993), "Empirical Cross-Section Dynamics in Economic Growth," *European Economic Review*, 37 (2/3): 426-434.

Sen, A. K. (2000), *Development as Freedom*, Anchor Books, New York, NY, 2000.

6.   APPENDIX

This appendix shows that the empirical frequency estimator used to estimate the transition probabilities in fact corresponds to the maximum likelihood estimator (MLE). Recall that, conditional on being in state i in the current period, $i=1,...,n$, there are n mutually exclusive states of the process next period, each with corresponding probabilities $p_{i1}, p_{i2},...,p_{in}, 0 \leq p_{ij} \leq 1, j=1,...,n, \sum_{j=1}^{n} p_{ij} = 1$. We are interested in estimating these transition probabilities based on observations on the N economies (or geographic states) that began in state i in the current period. In particular, the sample consists of count data on the number of economies in each state of the process:

(A1) $\{ x_{i1}, x_{i2},...,x_{in} \}$

where x_{ij} represents the number of economies that began in state i in the current period and moved to state j of the process next period. Thus:

(A2) $\sum_{j=1}^{n} x_{ij} = N$

In order to estimate these probabilities using the MLE, we first construct the likelihood function and then maximize it with respect to each of the probabilities. The estimator is non parametric since we know that, conditional on being in state i in the current period, the next state is the outcome of a draw from a multinomial distribution (we suppress the index i for convenience). To construct the likelihood function, we begin by recalling the multinomial probability mass function (PMF) with parameters $p_1, p_2,...,p_n, 0 \leq p_j \leq 1, j=1,...,n, \sum_{j=1}^{n} p_j = 1$:

(A3) $L(x_1, x_2,...,x_n, | p_1, p_2,...,p_n) = \dfrac{N!}{\prod_{j=1}^{n} x_j!} \prod_{j=1}^{n} p_k^{x_j}$

Assuming the observations are independently and identically distributed (i.i.d.), the maximum likelihood estimator based on the multinomial PMF is given by:

(A4) $p_j^{mle} = \arg_{pj} \max \left(\dfrac{N!}{\prod_{j=1}^{n} x_j!} \right) \prod_{j=1}^{n} p_k^{x_j}$

subject to

$$\sum_{j=1}^{n} x_j = N$$

(A5)
$$\sum_{j=1}^{n} p_j = 1$$

To find the maximum likelihood estimator for this distribution we maximize the probability of observing a given sample with respect to the parameters $p_1, p_2, \ldots, p_n$ by finding its partial derivatives with respect to $p_1, p_2, \ldots, p_n$ and equating them to zero:

$$\frac{\partial L(x_1, x_2, \ldots, x_{n,n} \mid p_1, p_2, \ldots, p_n)}{\partial p_k} = 0$$

(A6)
$$\left(\frac{N!}{\prod_{j=1}^{n} x_j!} \right) \prod_{j, j \neq k} p_j^{x_j} \; x_k \; p_k^{x_k - 1} = 0 \qquad \forall k$$

$$\left(\frac{N!}{\prod_{j=1}^{n} x_j!} \right) \prod_{j=1}^{n} p_j^{x_j} \; \frac{x_k}{p_k} = 0 \qquad \forall k$$

The first order conditions and the constraints can be summarized as:

$$\frac{x_k}{p_k} = \frac{x_l}{p_l} \qquad \forall j, l$$

$$\sum_{j=1}^{n} x_j = N$$

(A7)
$$\sum_{j=1}^{n} p_j = 1$$

Rearrange the first order conditions and combine them with the first constraint to obtain:

$$p_k x_l = p_l x_k \qquad \forall k, l$$

(A8)
$$x_l p_k = p_l \left(N - \sum_{j, j \neq k} x_j \right)$$

$$x_l p_k = p_l N - p_t \sum_{j, j \neq k} x_j$$

We then combine them with the second constraint to obtain:

$$x_l \left(1 - \sum_{j, j \neq k} p_j \right) = p_l N - p_t \sum_{j \neq k}^{n} x_j$$

(A9)
$$x_l - x_t \sum_{j \neq k}^{n} p_j = p_l N - p_t \sum_{j \neq k}^{n} x_j$$

Finally, use the fact that $p_k x_l = p_l x_k$, $\forall k, l$ to obtain:

$$x_l = p_l N$$

(A10)
$$\hat{p}_l^{MLE} = \frac{x_l}{N} \quad \forall l$$

Thus, the maximum likelihood estimator of the probability that an observation will be in state l next period given that it was in a given state this period is equal to the number of observations that began in a give state and moved to state l divided by the number of observations that began in the given state.