Kiel Institute of World Economics
Duesternbrooker Weg 120
24105 Kiel (Germany)

Kiel Working Paper No. 1086

Markov or Not Markov – This Should Be a Question

by

Frank Bickenbach and Eckhardt Bode

December 2001

Markov or Not Markov – This Should Be a Question

Abstract:

Although it is well known that Markov process theory, frequently applied in the literature on income convergence, imposes some very restrictive assumptions upon the data generating process, these assumptions have generally been taken for granted so far. The present paper proposes, resp. recalls chi-square tests of the Markov property, of spatial independence, and of homogeneity across time and space to assess the reliability of estimated Markov transition matrices. As an illustration we show that the evolution of the income distribution across the 48 coterminous U.S. states from 1929 to 2000 clearly has not followed a Markov process.

Keywords: Convergence, Markov process, chi-square tests, U.S. regional growth

JEL classification: C12, O40, R11

Frank Bickenbach
Kiel Institute of World Economics
24100 Kiel, Germany
Telephone: +49/431/8814-274
Fax: +49/431/8814-500
E-mail: fbickenbach@ifw.uni-kiel.de

Eckhardt Bode
Kiel Institute of World Economics
24100 Kiel, Germany
Telephone: +49/431/8814-462
Fax: +49/431/8814-500
E-mail: ebode@ifw.uni-kiel.de

1. Introduction

Since the late 1980s the issue of convergence or divergence of per-capita income and productivity has received considerable public attention, and has been addressed in a multiplicity of scientific papers. Depending on the underlying concept of convergence (unconditional or conditional β-convergence, σ-convergence, stochastic convergence), the statistical method employed (descriptive statistics, econometric approaches for cross-section, time-series, or panel data, Markov chain, or stochastic kernel estimations), and the geographic scope of analysis (countries, regions in single or groups of countries), the conclusions vary widely, ranging from rapid convergence to club convergence, and divergence. De la Fuente (1997), Durlauf and Quah (1999), and Temple (1999) have provided excellent reviews of the vast literature.

Most empirical approaches are based on hypotheses about the processes of interest rather than just describing them in a positive analysis. Often, some sort of a law (a 'law of convergence', a 'law of motion') is postulated to be valid even beyond the respective time period under consideration. The supposed relevance for future developments certainly has contributed to the popularity of respective approaches in the scientific as well as in the public sphere, as compared to simple descriptive statistics like the coefficient of variation. A politician, e.g., worrying about whether poor regions within his country, or poor countries in the world, may actually run the risk of being caught in a poverty trap will be strongly interested in a prediction for the future rather than just a description of the past.

In standard convergence regressions, as proposed by Barro and Sala-i-Martin (1991), and Mankiw et al. (1992), neoclassical growth theory is used to derive the hypothesis that income levels tend to converge. Having identified empirically a tendency towards (β-) convergence in the past, the underlying theoretical model suggests that convergence will continue until all regions will have the same per-capita income level (unconditional β-convergence) or, at least, an income level representing their specific behavioral and technical conditions (conditional β-convergence).

In Markov-chain approaches, as proposed by Quah (1993a; 1993b), the 'law of motion' driving the evolution of the income distribution is usually assumed to be memoryless and time-invariant. Having estimated probabilities of moving up or down the income hierarchy during a transition period of given length a stationary income distribution is calculated which characterizes the distribution the whole system tends to converge to over time. Although several authors (such as Quah himself, or Rey 2001b) emphasize that the stationary distribution represents merely a thought experiment it is often necessary to clarify the direction of the evolution since the estimated transition probability matrix by itself is not really informative about the evolution of the income distribution.[1]

The power of convergence regressions with respect to both describing comparative income growth processes in the period of analysis, and assessing the validity of neoclassical growth theory has been discussed extensively in the literature. Quah (1993a), and Durlauf and Quah (1999), e.g., have seriously challenged these approaches for several reasons. One reason is that the regression parameter of interest is biased towards convergence due to Galton's fallacy. Another reason is that convergence regressions cannot discriminate between neoclassical growth theory and alternative theoretical approaches, some of which having completely different implications. As a consequence, it may be useful to refrain from identifying the 'law of convergence', and from making inferences about the future on that basis. Just describing what happened in the past by switching to the concept of σ-convergence may be more appropriate. The evolution of the standard deviation, or of the coefficient of variation, is a reliable, unbiased indicator of convergence during the period of interest (Friedman 1992), provided the income distribution under consideration is normal, which can be tested for.

The power of the Markov chain approach, by contrast, has not yet been debated seriously.[2] The underlying statistical assumptions, namely the Markov property and time-invariance have just been taken for granted in empirical investigations so far. This is all the more surprising as the assumptions are quite restrictive, and as appropriate statistical tests are available in principle. The present paper will recall and illustrate a few test statistics that allow for assessing the reliability of the estimates and, in particular, of the stationary income distribution. Section 2 briefly sketches the Markov chain approach, and discusses relevant tests of the Markov property, of spatial independence, and of homogeneity of the estimated transition probabilities across space and time. Section 3 illustrates the tests by analyzing the evolution of the income distribution across the 48 coterminous U.S. states from 1929 to 2000. Section 4 concludes.

---

1 See, e.g., Quah (1996a); (1996b); Neven and Gouyette (1995); Fingleton (1997); (1999); Bode (1998a); (1998b); Magrini (1999); Rey (2001b); Bulli (2001).

2 Exceptions are Magrini (1999) and Bulli (2001).

2. The Markov chain approach

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 (see, e.g., Kemeny and Snell 1976: 24 ff.):

$$\begin{aligned} & P\{X(t+1)=j \mid X(0)=i_0, \ldots, X(t-1)=i_{t-1}, X(t)=i\} \\ & = P\{X(t+1)=j|X(t)=i\} \qquad (1) \\ & = p_{ij}. \end{aligned}$$

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

$$\Pi = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}, \qquad p_{ij} \geq 0, \ \sum_{j=1}^{N} p_{ij} = 1, \qquad (2)$$

which summarizes all N^2 transition probabilities p_{ij} (i, j = 1, …, N), and an initial distribution $\mathbf{h_0} = (h_{10}\ h_{20} \ldots h_{N0})$, $\Sigma_j h_{j0}=1$, describing the starting probabilities of the various states.

For illustration, let X be regional relative per-capita income, defined as $y_{rt} = Y_{rt} /[(1/R)\Sigma_r Y_{rt}]$ for region r and period t (r = 1, …, R; t = 0, …, T).[3] Divide the whole range of relative per-capita income into N disjunctive relative income classes (states). Then, a Markov transition probability is defined as the probability p_{ij} that a region is a member of income class j at t+1, provided it was in class i at t. The second row of the transition matrix (2), e.g., reports the probabilities that a member of the second-lowest income class (i=2) will descend into the lowest income class during one transition period (p_{21}), stay in the same class (p_{22}), change into the next higher income class (p_{23}), move upward two classes (p_{24}), and so on. Once having moved to another income class a region will behave according to the probability distribution relevant for that class. The initial probability vector $\mathbf{h_0}$, finally, describes the regional income distribution at the beginning of the first transition period, starting at t=0.

---

[3] The normalization by the national average is to control for global trends and shocks.

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. $\mathbf{h}_{t+m}=\mathbf{h}_t\mathbf{P}^m$. Moreover, if the Markov chain is regular the distribution converges towards a stationary[4] income distribution **h*** which is independent of the initial income distribution **h** ($\lim_{m\to\infty} \mathbf{h?}^m = \mathbf{h*}$). Comparing the initial income distribution ($\mathbf{h_0}$) 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.

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

$$\ln L = \sum_{i,j=1}^{N} n_{ij} \ln p_{ij} \qquad \text{s.t. } \Sigma_j\, p_{ij} = 1,\ p_{ij} \geq 0 \tag{3}$$

with respect to p_{ij} gives

$$\hat{p}_{ij} = n_{ij} / \sum\nolimits_j n_{ij} \tag{4}$$

as the asymptotically unbiased and normally distributed Maximum Likelihood estimator of p_{ij} (see, e.g., Anderson and Goodman 1957: 92; Basawa and Prakasa Rao 1980: 54 f.).[5] The standard deviation of the estimators can be estimated as (Bode 1998b)

$$\hat{s}_{\hat{p}_{ij}} = \left(\hat{p}_{ij}(1-\hat{p}_{ij})/n_i\right)^{1/2}. \tag{5}$$

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 $\hat{p}_{ij}$ are not allowed to be interpreted as Markov transition probabilities, and cannot be used to derive a stationary distribution.

4 In the literature, 'ergodic', or 'limiting' are used as synonyms for 'stationary'.

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

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 what follows we will concentrate on some of those assumptions of Markov process theory which are statistically testable. We will not deal with problems of inappropriate discretization of the income distribution which are discussed in Magrini (1999) and Bulli (2001).[6]

In practice, the estimation of Markov chains is subject 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, in principle, be 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 the regions, the higher the intensity of interaction, and thus the correlation of income levels, between neighboring regions tends to be. On the other hand, extending the geographical coverage of the sample increases the danger of lumping together regions whose development patterns are heterogeneous. Single regions, or certain groups of regions (like the southern states of the U.S.) may follow development paths that are different from the paths of other regions.

Likewise, the longer the time period under consideration, the higher the risk of structural breaks, i.e. regime changes which seriously affect the evolution of the income distribution.[7] 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.

---

6 Magrini (1999) and Bulli (2001) have argued that the usual ad-hoc discretization of the underlying continuous income distribution will probably remove the Markov property of the process. The crucial property of a Markov process, namely that future developments during any transition period t to t+1 do not depend on anything else but the own starting value at t, will be violated. As a result, the estimated probabilities cannot be interpreted as Markov transition probabilities, and the stationary distribution will be misleading.

7 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.

2. Some test statistics

The late 1950s and early 1960s witnessed a growing interest in the concept of Markov chains. A considerable number of journal articles and books dealing with test statistics for Markov chains were published (e.g. Anderson and Goodman 1957; Goodman 1958; Billingsley 1961a; 1961b; see also Basawa and Prakasa Rao 1980). Most prominently, chi-square, and Likelihood-Ratio tests were discussed. Both compare transition probabilities estimated simultaneously from the entire sample to those estimated from sub-samples obtained by dividing the entire sample into at least two mutually independent groups of observations. The criteria according to which the sub-samples are defined depend on the hypothesis to be tested against. Taken literally, the tests just compare multinomial distributions (rows of transition matrices) rather than Markov processes. A test of, e.g., whether two sub-samples ($r = 1, 2$) follow the same Markov process does not take into account whether or not the initial distributions ($\mathbf{h_{0r}}$) are likely to emerge from that Markov process.

The present paper will focus on the chi-square test; the LR test statistic is asymptotically equivalent. For details on the LR tests, see Anderson and Goodman 1957: 106 ff.; Kullback et al. 1962.

1. Tests for the entire transition matrix

There are several properties of a Markov process that can be tested for in the context of a data set pooled across several periods of time and several regions.

First, homogeneity over time (time-stationarity) can be checked by dividing the entire sample into T periods, and testing whether or not the transition matrices estimated from each of the T sub-samples differ significantly from the matrix estimated from the entire sample. More specifically, it tests H_0: $\forall t$: $p_{ij}(t)=p_{ij}$ ($t = 1, \ldots, T$) against the alternative of transition probabilities differing between periods: H_a: $\exists t$: $p_{ij}(t) \neq p_{ij}$. The chi-square statistic reads[8]

$$Q^{(T)} = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j\in B_i} n_i(t)\frac{\left(\hat{p}_{ij}(t)-\hat{p}_{ij}\right)^2}{\hat{p}_{ij}} \sim \text{asy } \chi^2\left(\sum_{i=1}^{N}(a_i-1)(b_i-1)\right), \qquad (6)$$

8 It is assumed that in each row (i) of the transition matrix for the entire sample there are at least two non-zero transition probabilities, and that the number of observations is positive for each of the T sub-samples.

where $\hat{p}_{ij}$ denotes the probability of transition from the i-th to the j-th class estimated from the entire sample (pooled across all T periods), and $\hat{p}_{ij}(t)$ the corresponding transition probability estimated from the t-th sub-sample. Since the $\hat{p}_{ij}(t)$ are assumed to be mutually independent across sub-samples under the H_0, the N^2 parameters can be estimated similar to (4) as $\hat{p}_{ij}(t) = n_{ij}(t)/n_i(t)$. $n_i(t)$ denotes the absolute number of observations initially falling into the i-th class within the t-th sub-sample. Only those transition probabilities are taken into account which are positive in the entire sample, i.e. $B_i = \{j: \hat{p}_{ij} > 0\}$; transitions for which no observations are available in the entire sample are excluded. Note that $n_i(t)$ may be zero: rows (i) for which no observations are available within a sub-sample do not contribute to the test statistic.

$Q^{(T)}$ has an asymptotic chi-square distribution with degrees of freedom equal to the number of summands in $Q^{(T)}$, except those where $n_i(t)=0$, minus the number of estimated transition probabilities $\hat{p}_{ij}$, both corrected for the number of restrictions ($\Sigma_j p_{ij}(t)=1$ and $\Sigma_j p_{ij}=1$). Consequently, the degrees of freedom can be calculated as $\Sigma_i a_i(b_i-1)-(b_i-1)$ where b_i ($b_i = |B_i|$)[9] is the number of positive entries in the i-th row of the matrix for the entire sample, and a_i is the number of sub-samples (t) in which observations for the i-th row are available ($a_i = |A_i|$; $A_i = \{t: n_i(t)>0\}$).

Second, homogeneity in the spatial dimension, implying H_0: $\forall r$: $p_{ij}(r)=p_{ij}$ ($r = 1, \ldots, R$) can be tested against the H_a of transition probabilities varying across regions, i.e. H_a: $\exists r$: $p_{ij}(r) \neq p_{ij}$, by

$$Q^{(R)} = \sum_{r=1}^{R} \sum_{i=1}^{N} \sum_{j \in B_i} n_i(r) \frac{\left(\hat{p}_{ij}(r) - \hat{p}_{ij}\right)^2}{\hat{p}_{ij}} \sim \text{asy}\ \chi^2 \left(\sum_{i=1}^{N} (c_i - 1)(b_i - 1) \right) \qquad (7)$$

where $c_i = |C_i|$; $C_i = \{r: n_i(r)>0\}$.

Third, the Markov property can be addressed directly by testing whether the process under consideration is memoryless, i.e. whether or not the transition probabilities are independent of the state k ($k = 1, \ldots, N$) a region was in at time t-1.

Fourth, and methodically quite similar, it can be tested whether the transition probabilities are independent across space, i.e., whether or not the transition probabilities are independent of the state s ($s = 1, ..., S$) a region's *neighboring* regions were in at time t.

---

9 $b_i = |B_i|$ means: b_i is the number of elements in set B_i.

The test principles for both the test of the Markov property, and that of spatial independence are similar to those sketched in eq. (6) and (7) above. For tests of spatial independence sub-samples are defined as in the concept of spatial Markov chains proposed by Rey (2001b). Given the definition of states *i* which divide the sample into *N* classes according to the regions' own income levels at *t*, Rey has suggested to define an additional set of states *s* for (average) relative income in neighboring regions at *t*, as illustrated in Figure 1. All regions with poor neighbors, e.g., constitute one sub-sample (*s*=1); those with medium-income neighbors a second, and those with rich neighbors a third one.

In the same way, the Markov property can be tested for by defining as additional states income classes the regions were in at time *t*-1: Regions that were poor at *t*-1 are allocated to the first sub-sample (*k*=1), those with median income to the second, and so forth.

Under the H_0 of time, resp. spatial independence, implying, $\forall k$: $p_{ij|k}=p_{ij}$, resp. $\forall s$: $p_{ij|s}=p_{ij}$, the transition matrices for all *N*, or *S* sub-samples can be estimated jointly because they are expected to be identical irrespective of the initial distribution of regions among the different sub-samples; the estimators are relative frequencies, similar to (4). The appropriate chi-square test statistic for time-independence is similar to (6) (just replace *t* and *T* by *k* and *N*), the test statistic for spatial independence is similar to (7) (replace *r* and *R* by *s* and *S*).

Figure 1 — Concept of spatial Markov chains by Rey (2001b)

income class neighbors (*s*)	initial distribution	transition matrices
s=1 (poor neighbors)	$h_{1\|1}$ (poor regions) … $h_{N\|1}$ (rich regions)	$p_{11\|1}$ … $p_{1N\|1}$ … $p_{N1\|1}$ … $p_{NN\|1}$
…	…	…
s=*S* (rich neighbors)	$h_{1\|S}$ (poor regions) … $h_{N\|S}$ (rich regions)	$p_{11\|S}$ … $p_{1N\|S}$ … $p_{N1\|S}$ … $p_{NN\|S}$

2. Tests for single states

The chi-square test statistics discussed above are quite flexible in that they can also be used to test whether or not a single state (i) in the overall sample (i-th row of the transition matrix for the entire sample) violates the underlying assumptions. Since the transition probabilities are assumed to be asymptotically independent across states under the H_0, define all observations in the i-th state to constitute an independent sample of its own, and perform the tests just introduced for this sample only. Homogeneity over time of the i-th state, implying H_0: $\forall t$: $p_{j|i}(t)=p_{j|i}$ ($t = 1, \ldots, T$), can be tested against non-stationarity (H_a: $\exists t$: $p_{j|i}(t) \neq p_{j|i}$) by (Anderson and Goodman 1957: 98)

$$Q_i^{(T)} = \sum_{t \in D_i} \sum_{j \in B_i} n_i(t) \frac{\left(\hat{p}_{ij}(t) - \hat{p}_{ij}\right)^2}{\hat{p}_{ij}} \sim \text{asy } \chi^2((d_i\text{-}1)(b_i\text{-}1)) \qquad (8)$$

where $D_i = \{t: n_i(t)>0\}$, $d_i = |D_i|$, and, as above, $b_i = |B_i|$, $B_i = \{j: \hat{p}_{ij} >0\}$.

Similarly, a test of spatial homogeneity of a single state i, i.e., H_0: $\forall r$: $p_{j|i}(r)=p_{j|i}$ ($r = 1, \ldots, R$) against H_a: $\exists r$: $p_{j|i}(r) \neq p_{j|i}$, is

$$Q_i^{(R)} = \sum_{r \in E_i} \sum_{j \in B_i} n_i(r) \frac{\left(\hat{p}_{ij}(r) - \hat{p}_{ij}\right)^2}{\hat{p}_{ij}} \sim \text{asy } \chi^2((e_i\text{-}1)(b_i\text{-}1)). \qquad (9)$$

In (9), $E_i = \{r: n_i(r)>0\}$, and $e_i = |E_i|$.

Note that (8) is similar to (6), and (9) is similar to (7), the only difference being that (8) and (9) compare only single rows in the transition matrices for all subsamples to the corresponding row in the matrix for the entire sample, while (6) and (7) compare whole matrices. Consequently, the statistics Q from (6) and (7) can be derived from (8) and (9) simply by summing up the Q_i across all states, i.e., $Q^{(T)} = \sum_i Q_i^{(T)}$, and $Q^{(R)} = \sum_i Q_i^{(R)}$.

(8) and (9) can also be applied to test for the Markov property, and for spatial independence; again, just a few indices have to be replaced. (8) can be used to test, e.g., the hypothesis that all regions that were poor at the beginning of the transition period under consideration (t to t+1) behave similarly irrespective of their income level in the past (at t-1). And (9) can be used to test, e.g., the hypothesis that all poor regions behave similarly irrespective of the income level of their neighbors at t.

3. Tests for single sub-samples

In some cases one might be interested in performing even more detailed tests comparing single sub-samples to the entire sample. For example, one might want to know whether or not a specific period differs significantly from the pattern estimated for the entire time span, or whether or not a specific region has evolved in line with the overall pattern. Such tests can be performed by using the chi-square test statistics (6) and (7) for a comparison of just two sub-samples (T=2, or R=2), namely the sub-sample of interest (t, or r) and the pool of the remaining observations in the entire sample. Since all sub-samples are assumed to be independent of each other, and to have the same distribution under H_0, any sub-sample may be isolated from the entire set of observations in this way.

Likewise, it can be tested whether or not a single state (i) within a single sub-sample (the t-th or r-th) differs significantly from the corresponding state estimated from the entire sample. This just requires defining all observations within the i-th state to constitute an independent sample of its own, split up this sample into two sub-samples (e.g., t and the rest), and compare both of them using (8) or (9).

4. Tests for a specified transition matrix

Finally, one may test whether or not the estimated transition matrix is equal to an exogenously given transition matrix, i.e., whether or not $p_{ij} = p_{ij}^0$ holds for all i,j = 1, …, N. The appropriate test statistic, known as χ^2 test of goodness of fit (Cochran 1952; Anderson and Goodman 1957: 96 f.), reads

$$Q^* = \sum_{i=1}^{N} \sum_{j \in F_i} n_i \frac{\left(p_{ij} - p_{ij}^0\right)^2}{p_{ij}^0} \qquad \sim \text{asy } \chi^2\left(\textstyle\sum_i^N (f_i - 1)\right). \tag{10}$$

$F_i = \{j\text{: } p_{ij}^0 > 0\}$ and $f_i = |F_i|$, i.e., the test is done only for those transition probabilities that are positive under the H_0.

For all the tests discussed above to be sufficiently exact, the definition of sub-samples in the time resp. the spatial dimension must be such that the numbers of observations from which the transition probabilities are estimated are sufficiently high to allow for reliable estimates (Cochran 1952). If the entire sample is quite small relative to the number of classes i, it does not leave too much room for defining additional sub-samples. Likewise, one cannot expect reliable results from testing whether or not a single row within a single sub-sample differs from the rest if there are only a few observations ($n_i(t)$) available for estimating the transition probabilities in this row.

3. Convergence among U.S. states 1929-2000

To illustrate the above-mentioned tests we use a data set of relative per-capita income pooled across the 48 coterminous U.S. states and 71 annual transition periods from 1929-1930 to 1999-2000.[10]

We arbitrarily divide the entire sample (3 408 observations) into five income classes with equal frequencies (quintiles) in order to ensure the number of observations per class to be sufficiently high to obtain reasonable estimates.[11] Table 1 gives the estimated (5x5) transition probability matrix and the stationary distribution obtained for the entire sample. Since the stationary income distribution shows somewhat higher probabilities in income classes around the median and lower probabilities in the extreme income classes than the initial distribution the estimates may be interpreted as reflecting convergence, if, indeed, the process under consideration is Markovian.

Table 1 — Estimated transition matrix for 48 U.S. states 1929-2000, annual transitions

income class	upper bound	initial distribution absolute	relative	transition probabilities (t to t+1) 1	2	3	4	5
1	0.82951	681	0.2	0.915	0.078	0.006	0.001	0
2	0.94741	682	0.2	0.065	0.828	0.103	0.003	0.001
3	1.03740	682	0.2	0.004	0.095	0.798	0.100	0.003
4	1.15897	682	0.2	0	0.010	0.100	0.837	0.053
5	∞	681	0.2	0	0	0	0.068	0.932
stationary distribution				0.172	0.212	0.219	0.215	0.182

Source: BEA, Regional Accounts Data; own estimation.

[10] Relative per-capita income is calculated as per-capita State Personal Income at current prices, divided by the unweighted average across all 48 coterminous U.S. states. The data is from the Bureau of Economic Analysis, Regional Accounts Data, released September 24, 2001 (http://www.bea.doc.gov/bea/regional/spi/).

[11] Note that the bounds of classes are fixed across the entire time-span under consideration.

1. Tests of homogeneity in time

To test for time-homogeneity we divide the 71 transition periods into 14 intervals (periods) of five annual transitions each. That is, we estimate 14 different transition probability matrices (T=14), each based on (5*48=) 240 observations,[12] in order to compare them simultaneously to the matrix for the entire time span (see Table 1). Using the test statistic (6) above[13] we obtain Q=365.3, which clearly rejects the H_0 of time-homogeneity (prob<0.01, 195 degrees of freedom). That is, the transition probabilities for the 48 U.S. states differ significantly over time; pooling over the entire time span of 71 transition periods is not appropriate.

There may, however, be one or more epochs in which the transition probabilities can be assumed to be constant. If there is no a priori information on the temporal location of major structural breaks that may have affected the evolution of the income distribution significantly, it may be useful to separately compare each of our 14 periods to the matrix for the whole time span using the test statistic (6) for T=2, as discussed in Section 2.2.3. First, we define the first period (1929-35) as one sub-sample, the remaining 13 periods (1935-2000) as a second one, and compare both to the entire sample. Afterwards, the second period (1935-40) is separated from the rest (1929-35, 1940-2000), and so on. The probability (prob-) values for the resulting 14 chi-square test statistics are plotted against the respective first years of the 14 periods in Figure 2. They show that significant deviations from the transition matrix for the entire sample concentrate on the years before 1950, and on the late 1990s.

To check whether at least 1950-1995 can be assumed to form a homogeneous sample we re-estimate the whole transition matrix for the reduced sample (Table 2), and test again for homogeneity over time. The resulting test statistic obtained from (6) is Q=112.0 which, at 96 degrees of freedom, does not indicate statistically significant differences between the transition matrices for the entire sample and the 9 periods of 5 years' length (prob=0.126). Consequently, the sample of 48 U.S. states may be pooled over the 45 annual transitions from 1950 to 1995, but not over a longer time span since structural breaks obviously occurred in the aftermath of World War II, and in the second half of the 1990s. While the former is well-documented in the literature[14] the latter should be taken

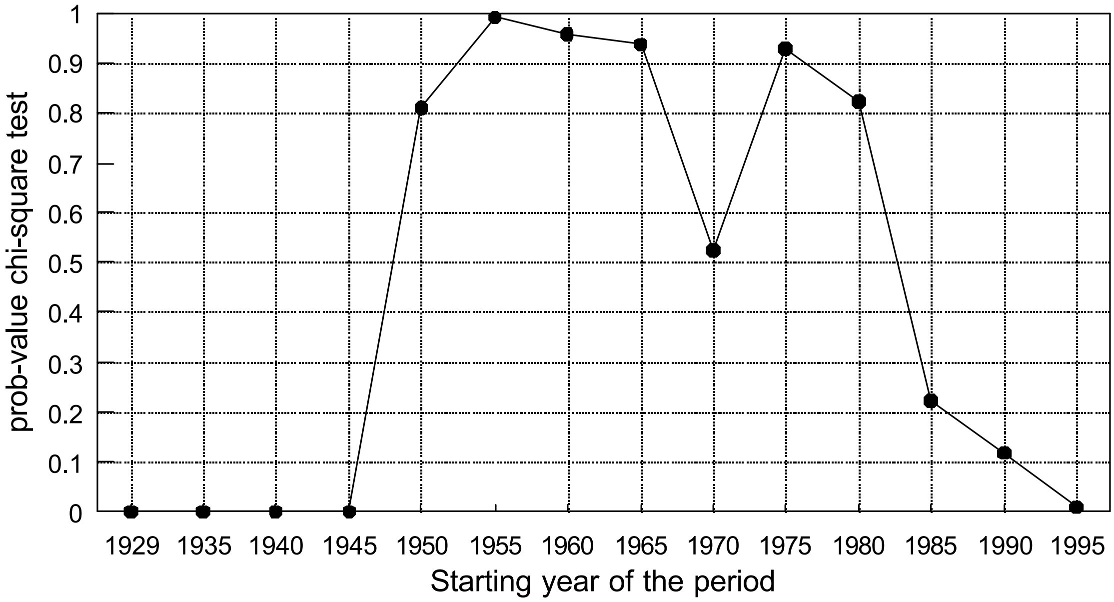
[12] The first period comprises 6 transition periods (1929-35) and 288 observations.

[13] The LR test is not applicable because some of the transition probabilities are zero in the temporal matrices but positive in the overall transition matrix.

[14] See, e.g., Carlino and Mills (1993), and Loewy and Papell (1996), who have identified structural breaks in the 1940s using unit-root tests.

with greater care since per-capita income figures for the 1990s are still based on interim estimates.

Figure 2 – Prob-values of chi-square tests of homogeneity over time – 48 U.S. states 1929-2000, annual transitions, divided into periods of 5-years



Source: BEA, Regional Accounts Data; own estimation.

Table 2 — Estimated transition matrix for 48 U.S. states 1950-1995, annual transitions

income class	upper bound	initial distribution absolute	initial distribution relative	transition probabilities (t to t+1) 1	2	3	4	5
1	0.85552	432	0.2	0.907	0.088	0.005	0	0
2	0.95438	432	0.2	0.074	0.838	0.088	0	0
3	1.03740	432	0.2	0.002	0.081	0.824	0.090	0.002
4	1.13509	432	0.2	0	0.005	0.100	0.859	0.037
5	∞	432	0.2	0	0	0	0.058	0.942

Source: BEA, Regional Accounts Data; own estimation.

2. Tests of spatial homogeneity

Tests of homogeneity in the spatial dimension based on the test statistic (7) can be illustrated by comparing transition matrices for different regions to that for the entire sample. While the H_0: $\forall r$: $p_{ij}(r)=p_{ij}$ is straightforward the H_a requires an exact specification of the potential spatial structure of heterogeneity. Several plausible sources come into mind: Each U.S. state may follow its own Markov process independent of other states. Or several states may constitute homogeneous spatial clusters, e.g., because they share common locational advantages and disadvantages, but different clusters may follow different Markov processes.

Since 45 observations per U.S. state (annual transitions 1950-1995) are not sufficient to estimate up to 25 transition probabilities reliably, we will concentrate on testing against an alternative of the second type, assuming the 8 BEA regions[15] to be independent sub-samples. The transition matrices estimated for the entire sample (same as in Table 2), and for each of the 8 BEA regions, as well as the statistics of the tests discussed in Sections 2.2.1. to 2.2.3 are reported in Table 3.

First, we compare all 8 BEA region-specific transition matrices to that for the entire sample. The respective chi-square statistic calculated according to (7) gives Q=338 which indicates that there are significant differences between the BEA regions (prob<0.01, 73 degrees of freedom).

Second, using eq. (9) in Section 2.2.2 we test for spatial homogeneity of single rows within the matrix for the entire sample, asking whether or not a single income class behaves similarly across BEA regions. The results can be found in the north-eastern corner of Table 3 (labeled "test of homogeneity" for the "entire sample"): The test hypothesis that BEA regions behave similarly within an income class (H_0: $\forall r$: $p_{j|i}(r)= p_{j|i}$) is clearly rejected for all five classes with very low error probabilities (prob<0.01).

Third, we compare the transition matrix for each BEA region to that of the entire sample by pooling the respective other 7 BEA regions into a second subsample, as has been described in the first paragraph of Section 2.2.3. The test statistics which are similar to (7), assuming R=2, are given in the rows labeled "whole matrix" at the bottom of each BEA region-specific section in Table 3. Only in the Great Lakes, and the Far West region the per-capita income distribution does evolve, by and large, in line with the entire sample; the error prob

[15] For a detailed definition see the Bureau of Economic Analysis at http://www.bea.doc.gov/bea/regional/docs/regions.htm.

Table 3 — Tests of spatial homogeneity across BEA regions, 48 U.S. states 1950-1995, annual transitions

income class	No of obs	transition probabilities 1	2	3	4	5	test of homogeneity d.o.f.	Q_i, Q	prob
				entire sample					
1	432	0.907	0.088	0.005	0	0	8	111.84	0.00
2	432	0.074	0.838	0.088	0	0	10	44.79	0.00
3	432	0.002	0.081	0.824	0.090	0.002	28	57.52	0.00
4	432	0	0.005	0.100	0.859	0.037	21	54.85	0.00
5	432	0	0	0	0.058	0.942	6	68.94	0.00
						whole matrix	73	337.95	0.00
				New England					
1	9	0.556	0.444	0	0	0	2	14.58	0.00
2	73	0.027	0.918	0.055	0	0	2	4.35	0.11
3	44	0	0.068	0.841	0.091	0	4	0.34	0.99
4	59	0	0	0.051	0.915	0.034	3	2.22	0.53
5	85	0	0	0	0.024	0.976	1	2.29	0.13
						whole matrix	12	23.77	0.02
				Mideast					
1	0	0	0	0	0	0	–	–	–
2	0	0	0	0	0	0	–	–	–
3	2	0	0	0	1	0	4	20.25	0.00
4	53	0	0	0.038	0.887	0.075	3	5.06	0.17
5	170	0	0	0	0.029	0.971	1	4.16	0.04
						whole matrix	8	29.47	0.00
				Great Lakes					
1	0	0	0	0	0	0	–	–	–
2	6	0	0.667	0.333	0	0	2	4.83	0.09
3	60	0	0.033	0.867	0.100	0	4	2.50	0.64
4	99	0	0	0.091	0.859	0.051	3	1.32	0.72
5	60	0	0	0	0.100	0.900	1	2.27	0.13
						whole matrix	10	10.92	0.36
				Plains					
1	28	0.464	0.464	0.071	0	0	2	83.54	0.00
2	59	0.254	0.644	0.102	0	0	2	33.16	0.00
3	165	0.006	0.042	0.848	0.097	0.006	4	8.47	0.08
4	62	0	0.032	0.258	0.710	0	3	34.77	0.00
5	1	0	0	0	1.000	0	1	16.32	0.00
						whole matrix	12	176.25	0.00

to be continued

Table 3 continued

income class	No of obs	transition probabilities 1	2	3	4	5	test of homogeneity d.o.f.	Q_i, Q	prob
				Southeast					
1	345	0.965	0.035	0	0	0	2	69.32	0.00
2	120	0.042	0.900	0.058	0	0	2	4.78	0.09
3	43	0	0.047	0.860	0.093	0	4	1.00	0.91
4	31	0	0	0.097	0.871	0.032	3	0.18	0.98
5	1	0	0	0	1.000	0	1	16.32	0.00
						whole matrix	12	34.76	0.01
				Southwest					
1	27	0.815	0.185	0	0	0	2	3.50	0.17
2	96	0.042	0.875	0.083	0	0	2	1.98	0.37
3	54	0	0.167	0.796	0.037	0	4	7.87	0.10
4	3	0	0	0.667	0.333	0	3	10.86	0.01
5	0	0	0	0	0	0	–	–	–
						whole matrix	11	24.21	0.01
				Rocky Mountains					
1	23	0.826	0.174	0	0	0	2	2.33	0.31
2	78	0.077	0.782	0.141	0	0	2	3.41	0.18
3	46	0	0.261	0.674	0.065	0	4	22.56	0.00
4	62	0	0	0.081	0.855	0.065	3	2.07	0.56
5	16	0	0	0	0.375	0.625	1	30.65	0.00
						whole matrix	12	61.02	0.00
				Far West					
1	0	0	0	0	0	0	–	–	–
2	0	0	0	0	0	0	–	–	–
3	18	0	0	0.889	0.111	0	4	1.79	0.77
4	63	0	0	0.048	0.952	0	3	5.82	0.12
5	99	0	0	0	0.040	0.960	1	0.72	0.40
						whole matrix	8	8.34	0.40

Source: BEA, Regional Accounts Data; own estimation.

abilities being 0.36 (Great Lakes) and 0.4 (Far West), respectively. For the other BEA regions, by contrast, the error probabilities are below 0.05, indicating that these regions are not well represented by the figures estimated for the U.S. as a whole.

And finally, we compare single income classes for single BEA regions to the corresponding income class estimated for the U.S. as a whole by proceeding as described in the second paragraph of Section 2.2.3. The test statistics reported in Table 3 to the right of the BEA region-specific transition matrices draw a fairly mixed picture.[16] For example, in the Rocky Mountains region it seems to be the regions with median, and with very high income (income classes 3 and 5) that behave differently from the U.S. average. For both classes an above-average probability of becoming poorer is obtained ($\hat{p}_{32}(Rocky)$=0.261, $\hat{p}_{54}(Rocky)$=0.375, compared to 0.081, and 0.058 for the entire sample).

3. Tests of the Markov property

As noted earlier the Markov property requires the transition probabilities from t to t+1 to depend only on a region's initial state at t but not on its state at t-1 (or any earlier point in time). This property can be tested against some sort of first-order serial autocorrelation, i.e. against the hypothesis that regions belonging to the same income class at t behave differently depending on their state at t-1. We define five sub-samples k = 1, ..., 5 for states the regions were in at t-1, such that $i(t\text{-}1) = k(t)$. E.g., regions that were in the first income class at t-1 are allocated to the first sub-sample (k=1), those that were in the second class are put into k=2, and so on. For each of these sub-samples we estimate a separate matrix from observed transitions from t to t+1 in the usual way.

The estimated transition matrices for the entire sample (income class at t-1 = 'all') and for the five sub-samples as well as the various test statistics are given in Table 4.[17] The general test comparing the matrices for all five sub-samples to that for the entire sample simultaneously (similar to eq. 6 in Section 2.2.1) produces Q=436.2 which, at 30 degrees of freedom, indicates extremely significant differences (prob<0.01). Consequently, the evolution of the income distribution across U.S. states cannot be assumed to be independent of the past. This is not only true for the entire sample but also for each of the five income classes, as the test statistics in the north-eastern part of Table 4 indicate. If the H_0 of time independence was true, the rows of the matrices for the 5 sub samples would be equal to the corresponding row of the matrix for the entire sample, and the tests would not indicate significant differences. This is clearly

[16] Of course, several of the test statistics are quite poorly reliable since the numbers of underlying observations are low.

[17] Note that the matrix for the entire sample ("income class at t-1 = 'all'") differs slightly from that in Table 2 because the first transition period (1950-51) is needed for the serial lag.

Table 4 — Tests of the Markov property, 48 U.S. states 1951-1995, annual transitions

income class at *t*-1	*t*	No of obs	transition probabilities (*t* to *t*+1) 1	2	3	4	5	test of Markov prop d.o.f.	Q_i, Q	prob
all	1	422	0.908	0.088	0.005	0	0	4	235.7	0.00
all	2	423	0.073	0.839	0.087	0	0	6	24.54	0.00
all	3	422	0.002	0.078	0.827	0.090	0.002	12	76.04	0.00
all	4	423	0	0.005	0.097	0.865	0.033	6	64.48	0.00
all	5	422	0	0	0	0.059	0.941	2	35.40	0.00
							whole matrix	30	436.2	0.00
1	1	391	0.928	0.072	0	0	0	2	43.43	0.00
1	2	37	0.243	0.730	0.027	0	0	2	18.25	0.00
1	3	2	0	0.500	0.500	0	0	4	5.02	0.29
1	4	0	0	0	0	0	0	0	–	–
1	5	0	0	0	0	0	0	0	–	–
							whole matrix	8	66.71	0.00
2	1	30	0.667	0.300	0.033	0	0	2	24.24	0.00
2	2	350	0.057	0.860	0.083	0	0	2	8.75	0.13
2	3	38	0	0.289	0.684	0	0.026	4	38.83	0.00
2	4	0	0	0	0	0	0	0	–	–
2	5	0	0	0	0	0	0	0	–	–
							whole matrix	8	71.82	0.00
3	1	1	0	0	1.000	0	0	2	210.5	0.00
3	2	34	0.059	0.735	0.206	0	0	2	6.51	0.05
3	3	340	0.003	0.059	0.868	0.071	0	4	23.64	0.33
3	4	38	0	0.053	0.342	0.605	0	3	50.80	0.00
3	5	1	0	0	0	1.000	0	1	15.92	0.00
							whole matrix	12	307.4	0.00
4	1	0	0	0	0	0	0	0	–	–
4	2	2	0	1.000	0	0	0	2	0.38	0.83
4	3	42	0	0.024	0.643	0.333	0	4	34.56	0.00
4	4	361	0	0	0.075	0.898	0.028	3	29.22	0.00
4	5	16	0	0	0	0.313	0.688	1	19.14	0.00
							whole matrix	10	83.30	0.00
5	1	0	0	0	0	0	0	0	–	–
5	2	0	0	0	0	0	0	0	–	–
5	3	0	0	0	0	0	0	0	–	–
5	4	24	0	0	0.042	0.792	0.167	3	14.80	0.00
5	5	405	0	0	0	0.047	0.953	1	27.42	0.00
							whole matrix	4	42.21	0.00

Source: BEA, Regional Accounts Data; own estimation.

not the case; there is not a single income class for which the previous income level is irrelevant. The following four examples may serve as an illustration:[18]

– Take, e.g., the first row in the matrix for the second sub-sample (income class at t-1=2, and at t=1) representing regions that descended from the second to the first income class just the period before (t-1 to t). These regions have a considerably higher probability of becoming richer again ($\hat{p}_{12|2}$=0.300) than regions that were poor before ($\hat{p}_{12|1}$=0.072), and than regions on average $\hat{p}_{12}$=0.088.

– Similarly, regions that just scaled up from the lowest to the second-lowest class have a considerably higher probability of falling back again than regions that have already been in the second class for a longer time; the respective probabilities being $\hat{p}_{21|1}$=0.243, $\hat{p}_{21|2}$=0.057, and $\hat{p}_{21}$=0.073.

– At the upper end of the income hierarchy, very rich regions that were very rich before tend to have a higher probability of staying very rich than very rich regions that were poorer before, i.e. $\hat{p}_{55|5}$=0.953 > $\hat{p}_{55|4}$=0.688.

– Finally, consider the third income class in t: The probability of staying in that class if a region was in there before ($\hat{p}_{33|3}$=0.868) is higher than both the probabilities of regions that were poorer, or richer before: $\hat{p}_{33|2}$=0.684, and $\hat{p}_{33|4}$=0.643.

Obviously, history matters a lot. At least some of the movements between classes are temporary; the probabilities estimated from the entire sample are rather poor predictors of the real behavior of regions, at least in several cases.

---

[18] The definition of classes for the tests of the Markov property has produced a very obvious outlier, namely North Dakota's transition from 1979 to 1980 and 1981. This is the only observation falling in the first class of the third sub-sample (income class at t-1=3). The per-capita income declined sharply from 1979 to 1980; South Dakota descended from the third to the lowest income class. In the next period, income rose again; the state returned to the third class. Since leaps of this kind across two class boundaries have been very rare among the poorest regions they are penalized strongly by the test statistic. Eliminating this outlying observation from the data set does not change the overall conclusions significantly, however.

4. Tests of spatial independence

Although the Markov approach requires stochastically independent observations income dynamics in one state may be affected by geographic spillovers from respective neighboring states. As Rey and Montouri (1999) and Rey (2001a), (2001b) have shown by means of several statistical indicators, spatial dependence among neighboring U.S. states is quite strong: There seem to be sort of spillovers among neighboring states such that income dynamics in one state is not independent of whether its neighbors are – on average – comparatively rich or poor. Similar results have been obtained for regions in various other countries (e.g. Fingleton and McCombie 1998, Bode 1998b, 2001, 2002) as well as at the international level (Keller 2000, Fingleton 2000).

For empirically illustrating the test of spatial independence proposed in Section 2 we distinguish 5 spatial Markov chains by (again arbitrarily) dividing the sample (48 states, 1950-1995, 2 160 observations) into 5 income classes $s = 1, \ldots, S$ for different income levels in neighboring states. An observation is allocated to s=1 if the average relative per-capita income at time t in the neighboring states falls into the first quintile across all observations, i.e. if the region-year under consideration is among the 432 observations (20 per cent of the entire sample) for region-years having the poorest neighbors. For each of the resulting spatial Markov chains we test the hypothesis that the transition probabilities are equal to the transition matrix in Table 2.

The results strongly support earlier findings: the whole system is not independent across space, the test statistic Q=144 being highly significant (prob <0.01; 47 d.o.f.; Table 5). For three out of five income classes of the entire sample the income level in a state's geographic neighborhood is important. The tests comparing single rows across all sub-samples (see Section 2.2.2., eq. 9) indicate that there are significant differences for states with low, above-median, and high income (classes 1, 4, and 5; north-eastern part of Table 5). And there is not a single among the five sub-samples for different income levels of neighbors that does not show significant differences to the entire sample, as the test statistic discussed in Section 2.2.3 indicates (rows labeled "whole matrix"). Obviously, e.g., a poor state has a substantially lower probability of becoming richer if its neighbors are poor as well ($\hat{p}_{12|1}$=0.032), compared to the average across all states which is estimated to be $\hat{p}_{12} + \hat{p}_{13}$=0.093. Similarly, a very rich state has a lower probability of becoming poorer if its neighbors are very rich as well ($\hat{p}_{54|5}$=0.024, compared to $\hat{p}_{54}$=0.058).

Table 5 — Tests of spatial dependence among immediate neighbors, 48 U.S. states 1950-1995, annual transitions

income class	No of obs	transition probabilities 1	2	3	4	5	test of homogeneity d.o.f.	Q_i, Q	prob
				entire sample					
1	432	0.907	0.088	0.005	0	0	8	48.17	0.00
2	432	0.074	0.838	0.088	0	0	8	13.84	0.09
3	432	0.002	0.081	0.824	0.090	0.002	16	19.21	0.26
4	432	0	0.005	0.100	0.859	0.037	12	25.70	0.01
5	432	0	0	0	0.058	0.942	3	37.09	0.00
						whole matrix	47	144.0	0.00
				poor neighbors (s=1)					
1	253	0.968	0.032	0	0	0	2	27.36	0.00
2	121	0.033	0.893	0.074	0	0	2	4.77	0.09
3	43	0	0.140	0.791	0.070	0	4	2.53	0.64
4	15	0	0	0.133	0.867	0	3	0.83	0.84
5	0	0	0	0	0	0	0	–	–
						whole matrix	11	35.48	0.00
				neighbors with below-median income (s=2)					
1	67	0.731	0.239	0.030	0	0	2	34.12	0.00
2	103	0.107	0.806	0.087	0	0	2	2.12	0.35
3	136	0	0.044	0.897	0.059	0	4	7.71	0.10
4	115	0	0	0.087	0.870	0.043	3	1.17	0.76
5	11	0	0	0	0.455	0.545	1	32.58	0.00
						whole matrix	12	77.70	0.00
				neighbors with median income (s=3)					
1	75	0.840	0.160	0	0	0	2	6.23	0.04
2	71	0.141	0.761	0.099	0	0	2	5.83	0.05
3	102	0.010	0.069	0.794	0.118	0.010	4	7.98	0.09
4	96	0	0	0.083	0.875	0.042	3	1.01	0.80
5	88	0	0	0	0.080	0.920	1	0.95	0.33
						whole matrix	12	21.99	0.04
				neighbors with above-median income (s=4)					
1	33	0.970	0.030	0	0	0	2	1.67	0.43
2	89	0.079	0.831	0.090	0	0	2	0.04	0.98
3	106	0	0.085	0.792	0.123	0	4	2.48	0.65
4	81	0	0.025	0.198	0.716	0.062	3	22.29	0.00
5	123	0	0	0	0.065	0.935	1	0.16	0.69
						whole matrix	12	26.64	0.01
				rich neighbors (s=5)					
1	4	0.750	0.250	0	0	0	2	1.33	0.51
2	48	0	0.896	0.104	0	0	2	4.38	0.11
3	45	0	0.156	0.778	0.067	0	4	4.12	0.39
4	125	0	0	0.056	0.928	0.016	3	7.25	0.06
5	210	0	0	0	0.024	0.976	1	8.70	0.00
						whole matrix	12	25.77	0.01

Source: BEA, Regional Accounts Data; own estimation.

5. Are these tests purely academic exercises?

To exemplify the sensitivity of empirical results to violations of the requirements of the method we will finally compare the stationary distribution calculated from the transition matrix for the entire sample (48 states, 1950-1995) to those from the transition matrices for single BEA regions (see Table 3). A similar discussion of the effects of spatial dependence on stationary distributions can be found in Rey (2001b). Note that the following exercises are purely illustrative.

A conventional interpretation of the stationary distribution for the U.S. as a whole (Table 6, first row), as has been adopted frequently in the literature, would conclude that there is some good news for states that are lagging behind, and some bad news for the leaders: Apparently, there is convergence among U.S. states. Compared to the initial distribution (0.2 in each class) the populations in the extreme classes have decreased.[19]

With a view to the BEA region-specific limiting distributions, this conclusion may be appropriate for the Plains, indeed, although the tendency towards concentrating at the median (of all 48 states) appears to be much stronger there.

Table 6 — Stationary income distributions calculated from estimated transition matrices for 8 BEA regions 1950-1995, annual transitions

BEA-region	income class				
	1	2	3	4	5
USA	0.186	0.225	0.236	0.210	0.144
New England	0.011	0.186	0.150	0.267	0.385
Mideast	0	0	0.010	0.277	0.712
Great Lakes	0	0.036	0.363	0.399	0.202
Plains	0.108	0.216	0.496	0.176	0.003
Southeast	0.255	0.213	0.267	0.257	0.008
Southwest	0.128	0.571	0.285	0.016	0
Rocky Mountain	0.177	0.401	0.217	0.175	0.030
Far West	0	0	0.300	0.700	0

Source: BEA, Regional Accounts Data; own estimation.

[19] Shaded cells in Table 6 indicate peaks of the distributions.

For the rest of the BEA regions, however, the general picture seems to be of very limited relevance. The figures suggest that New England, Mideast, and Far West states as well as those at the Great Lakes have little reason to worry about falling back to mediocrity. And the supposedly good news for the south and the Rocky Mountain states may be way too optimistic. It seems as if the majority of them continue to be comparatively poor. Of course, one has to bear in mind that there may have been a structural break in the 1990s, and that in the evolution of the income distributions *within* BEA regions may not be homogeneous, which has not been tested for.

4. Conclusions

Although Markov process theory offers a couple of desirable features for convergence analysis such as the possibility to determine a stationary income distribution, it requires some very restrictive assumptions to be met. Quite surprisingly, these assumptions have generally been taken for granted in the convergence literature so far. This is all the more surprising, as appropriate tests have been available since the late 1950s, and are quite simple to implement. The present paper has proposed, resp. recalled a number of tests to assess the properties of estimated Markov transition matrices.

In summary, these tests turn out to be useful tools. The chi-square statistic discussed in this paper is very flexible in use. It can be used for a wide variety of tests, ranging from tests of the Markov property and spatial dependence to homogeneity of observed processes over time and space. It can be used to compare whole systems of transition matrices as well as single rows in transition matrices for single sub-samples. All tests, however, require the number of observations to be large enough to allow for reasonably accurate estimates of transition probabilities.

As has been illustrated, the evolution of the income distribution across the 48 coterminous U.S. states from 1929 to 2000 clearly does not follow a Markov process. Rather, income growth has been autoregressive in both time, and space. Regional clusters of states apparently have followed different laws of motion (if any), and there has been a structural break in the aftermath of World War II, that has significantly affected the evolution of the income distribution. Another structural break may have occurred in the 1990s. These features should be taken into consideration when making inferences about the evolution of the regional income distribution in the U.S.

References

Anderson, T.W., L.A. Goodman (1957). Statistical Inference about Markov Chains. *Annals of Mathematical Statistics* 28 (1): 89–110.

Barro, R.J., X. Sala-i-Martin (1991). Convergence. *Journal of Political Economy* 100 (2): 223–251.

Basawa, I.Y., B.L.S. Prakasa Rao (1980). *Statistical Inference for Stochastic Processes*. London.

Billingsley, P. (1961a). *Statistical Inference for Markov Processes*. Chicago: University of Chicago Press.

Billingsley, P. (1961b). Statistical Methods in Markov Chains. *Annals of Mathematical Statistics* 32 (1): 12–40.

Bode, E. (1998a). Wirtschaftliche Konvergenz in Deutschland. In: B. Fischer, T. Straubhaar (eds.), *Ökonomische Konvergenz in Theorie und Praxis*. Baden-Baden: Nomos.

Bode, E. (1998b). *Lokale Wissensdiffusion und regionale Divergenz in Deutschland*. Kieler Studien 293. Tübingen: Mohr.

Bode, E. (2001). Is Regional Innovative Activity Path-dependent? An Empirical Analysis for Germany. Kiel Working Paper No. 1058. Kiel Institute of World Economics.

Bode, E. (2002). R&D, Localized Knowledge Spillovers, and Endogenous Regional Growth: Evidence from Germany. Forthcoming in L. Schätzl and J. Revilla Diez (eds.), *Technological Change and Regional Development in Europe*. Heidelberg: Physika.

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

Carlino, G.A., L.O. Mills (1993). Are U.S. Regional Incomes Converging? A Time Series Analysis. *Journal of Monetary Economics* 32 (2): 335–346.

Cochran, W.G. (1952). The χ^2 Test of Goodness of Fit. *Annals of Mathematical Statistics* 23: 315–345.

De la Fuente, A. (1997). The Empirics of Growth and Convergence: A Selective Review. *Journal of Economic Dynamics and Control* 21 (1): 23–73.

Durlauf, S.N., D.T. Quah (1999). The New Empirics of Economic Growth. In: Taylor, J.B. (ed.), *Handbook of Macroeconomics*, Vol. 1A. Amsterdam: Elsevier.

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.

— (1999). Estimates of Time to Economic Convergence: An Analysis of Regions of the European Union. *International Regional Science Review* 22 (1): 5–34.

— (2000). Convergence: International Comparisons Based on a Simultaneous Equation Model with Regional Effects. *International Review of Applied Economics* 14 (3): 285–305.

—, J.S.L. McCombie (1998). Increasing Returns and Economic Growth: Some Evidence for Manufacturing from the European Union Regions. *Oxford Economic Papers* 50 (1): 89–105.

Friedman, M. (1992). Do Old Fallacies Ever Die? *Journal of Economic Literature* 30 (4): 2129–2132.

Goodman, L.A. (1958). Simplified Runs Tests and Likelihood Ratio Tests for Markov Chains. *Biometrica* 45 (1/2): 181–197.

Keller, W. (2000). Do Trade Patterns and Technology Flows Affect Productivity Growth? *World Bank Economic Review* 14 (1): 17–47.

Kemeny, J.G., J.L. Snell (1976). *Finite Markov Chains*. Heidelberg: Springer.

Kullback, S., M. Kupperman, and H.H. Ku (1962). Tests for Contingency Tables and Markov Chains. *Technometrics* 4 (4): 573–608.

Loewy, M.B., D.H. Papell (1996). Are U.S. Regional Incomes Converging? Some Further Evidence. *Journal of Monetary Economics* 38 (3): 587–598.

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

Mankiw, N.G., D. Romer, D.N. Weil (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics* 107 (2): 407–437.

Neven, D., C. Gouyette (1995). Regional Convergence in the European Community. *Journal of Common Market Studies* 33 (1): 47–65.

Quah, D.T. (1993a). Galton's Fallacy and Tests of the Convergence Hypothesis. *The Scandinavian Journal of Economics* 95 (4): 427–443.

Quah, D.T. (1993b). Empirical Cross-section Dynamics in Economic Growth. *European Economic Review* 37 (2): 426–434.

Quah, D.T. (1996a). Regional Convergence Clusters Across Europe. *European Economic Review* 40 (3): 951–958.

Quah, D.T. (1996b). Empirics for Economic Growth and Convergence. *European Economic Review* 40 (6):1353–1375.

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.

— (2001b). Spatial Empirics for Economic Growth and Convergence. *Geographical Analysis* 33 (3): 195–214.

—, B.D. Montouri (1999). US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies* 33 (2): 143–156.

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature* 37(1): 112-156.