Working Paper No. 58/00

Distribution Dynamics: Stratification, Polarisation and Convergence Among OECD Economies, 1870-1992

Philip Epstein, Peter Howlett and Max-Stephan Schulze

Department of Economic History
London School of Economics 2000

1. Introduction

Since the 1980s the debate about economic convergence has dominated empirical work about the dynamics of growth.[1] Economic historians have been attracted, in particular, by stories of club convergence.[2] However, the analytical foundations of most of the work in this area have rested on linear, or more usually log-linear, regression analysis. Thus, the results tend to be dependent on a conditional average in which time is the dominant player.[3] This is surprising as space, and issues of distribution, have long been important to both theorists and historians. A notable exception to the 'regression school' has been the work on distribution dynamics pioneered in a series of papers by Danny Quah (1993, 1996, 1997). He believes that only by considering the issues of growth and distribution simultaneously can we understand their underlying dynamics. He has argued, for example, that there is no simple causal relationship between the concepts of β-convergence and σ-convergence and that similar stories of global (or club) convergence may be driven by very different stories of individual economy mobility. This is an approach that should appeal to economic historians (both because it can encompass a rich diversity of individual economy experience and because it emphasises that same diversity). We hope to illustrate this by considering the experience of some of the leading OECD economies since 1870 within an explicit distribution dynamics framework.[4]

---

[1] Some of the more influential papers in this area include Solow (1956), Baumol (1986), Romer (1986), Barro and Sala-I-Martin (1991), Mankiw, Romer and Weil (1992), Bernard and Durlauf (1995).

[2] Examples of those who have told explicit club convergence stories are Tortella (1994) and Toniolo (1998). An important proponent of the convergence approach is Williamson (1996). A summary of the literature can be found in Broadberry (1996). A notable exception to the club convergence approach is Mills and Crafts (2000).

[3] And time in regression analysis is not without its problems, as the voluminous work on unit roots, cointegration and the assessment of structural breaks is testimony to.

[4] We have explored these issues in another paper, Epstein, Howlett and Schulze (1999).

2. Scope and Data

We consider two groups of economies. The first represents 17 advanced OECD economies (in their post-World War II boundaries) for which homogenous, annual long run historical data are available: Australia, Austria, Belgium, Canada, Denmark, Finland, France, West Germany, Italy, Japan, Netherlands, New Zealand, Norway, Spain, Sweden, United Kingdom and the United States. GDP per capita levels in 1990 international dollars have been taken from the well known and widely used Maddison data set (1995, pp.194-201).[5] These data are continuous, cover the whole of the twentieth and much of the nineteenth century, and have been much used in the recent historical literature. They have been extended and revised using new GDP estimates for 19th Austria, which are based on computations for the former Habsburg Empire by Schulze (2000), and for Spain drawing on Prados de la Escosura (1995).[6] The gaps in the Maddison series for Japan have been closed, relying on the work of Ohkawa and his collaborators (Ohkawaw, Shinohara and Meisner, 1979; Japan Statistical Association,1988).

Economic historians often discuss long term growth in terms of regimes, epochs or phases. For example, Williamson (1996) in discussing the OECD club identifies three epochs: the late nineteenth century was characterised by fast growth, globalisation, and convergence; 1914-50 witnessed slow growth, de-globalisation, and divergence; and the post-1950 era has experienced fast growth, globalisation and convergence. This periodisation is similar to that proposed by Angus Maddison (1995, pp.59-87). Maddison argues that since 1820 there have been five distinct phases of development: 1820-70; 1870-1913; 1913-50; 1950-73; 1973-

[5] Our group of countries is the same as the group of 17 advanced economies defined by Maddison except that Switzerland (for which annual data is only available after 1900) is replaced by Spain.

[6] For the income ratios used to derive per capita income levels in the territories of modern-day Austria on the basis of estimates for imperial Austria see Good and Ma (1998).

92.[7] The 1870-1913 phase is, according to Maddison, characterised as 'a relatively peaceful and prosperous era' in which 'per capita growth accelerated in all regions and in most countries'. This phase of growth eventually gave way to 'an era deeply disturbed by war, depression, and beggar-your-neighbour policies…a bleak age, whose potential for accelerated growth was frustrated by a series of disasters'. The 1950-73 period was 'a golden age of unparalleled prosperity' in which income per head in all regions 'grew faster than in any other phase.' In the period after 1973 inflationary pressures, the breakdown of the Bretton Woods fixed exchange rate system, and the oil price shocks brought about 'a sharp reduction in the pace of economic growth throughout the world…and the momentum of the golden age' was lost.

Given the broad agreement between these scions of quantitative economic history, for analytical purposes the period 1870 to 1992 has been divided into three sub-periods: 1870-1913 (the pre-1914 period), 1914-1950 (the transwar period) and 1951-1992 (the post-war period). These periods thus conform to the regime shifts identified by Maddison and Williamson. By definition this is a biased grouping, representing 17 of the most developed nations in the world at the end of the twentieth century. If the theoretical literature is correct, the forces of convergence should be very strong among this particular group in the pre-1914 and the post-war periods.

Much of the current historiography (including Maddison) also tends to discuss the post-war period in terms of the 'Golden Age' (1951-1973) and post-'Golden Age' (1974-92). Data limitations mean that we cannot use the full range of techniques we wish to illustrate using just the 17 advanced countries. Thus to be able to investigate the Golden Age and post-Golden Age periods we have also considered a larger group of 24 economies. The additional economies are Czechoslovakia, Greece,

---

[7] He further sub-divides the period 1913-50 into four sub-periods (1913-29, 1929-38, 1938-44,

Hungary, Ireland, Portugal, Switzerland and Turkey. The addition of these 7 economies will increase the variance compared to the group of 17 advanced economies, as they are a poorer group of economies. For example, in 1992 the average real GDP per capita in the seven additional economies was 65% of the average level of real GDP per capita in the 17 advanced economies.[8]

For each year we standardised the observations to the *average* level of real GDP per head among the 17 (or 24) advanced economies in that year (with the average taking a value of 1.00).

3. Some Traditional Empirics

3.1. s-Convergence

The simplest measure of s-convergence (the reduction in the dispersion of income levels over time) is the coefficient of variation. For any given group of economies, s-convergence implies that over time the variation in their incomes relative to their mean income will decline. Thus convergence should be reflected in a decline of the coefficient of variation over time. Figure 1 shows this measure for the both sets of economies in order to provide some background for the discussion of distribution dynamics.

Four observations on the dispersion in *levels* of per capita incomes among the 17 can be made. First, during the late nineteenth century up to 1913 dispersion declined, after an initial increase in the 1870s. This decline occurred largely in the 1880s and 1890s and there was hardly any change in income dispersion in the later years. Second, the First World War and its immediate aftermath brought about a pronounced yet temporary rise in income dispersion, which gave way to further reduction in the 1920s and early 1930s, with some modest rise in dispersion in the

---

and 1944-49).

[8] In 1992, after the collapse of the communist regimes of Eastern Europe, this figure is reduced to 60%.

later 1930s. Third, during the Second World War the variance in per capita income across the 17 economies rose dramatically and well above the levels observed for the 1870s, only to decline again in the immediate post-war years. Thirdly, there was a virtually uninterrupted diminution of variance in incomes in the post-war period until the mid-1970s, after which they were relatively stable. It is also worth noting that the low dispersion levels of the mid-1930s were not again reached until the beginning of the 1960s.

This evidence would suggest that during the pre-World War I period a process of gradual convergence got under way that was temporarily reversed during the two world wars. It is well worth noting, though, that this process seemed to have received little further impetus in the very late 19th and early 20th centuries, i.e. during the years with generally faster economic growth than in the two preceding decades. Perhaps surprisingly, neither the protectionist trade regimes of the 1920s, nor the great depression and the rise of protectionist trade blocs in the 1930s appear to have made much of an adverse impact on this process. After the end of the Second World War, per capita income differentials among the 17 economies became progressively smaller and in this sense one can say that post-war growth was accompanied by convergence.

For the 24 economies there is also a sharp decline in income dispersion in the post-war period. However, this decline comes to an end in the mid-1970s and thereafter dispersion is stable or even rising. This provides some empirical justification in treating the Golden Age and Post-Golden Age as two separate regimes.

3.2. Empirical Distributions

In order to illustrate the importance of *mobility*, we consider briefly the *actual* empirical distributions associated with the 17 advanced OECD economies. The technique employed here allows us to identify five income states (state 1 representing the lowest level and state 5 the highest level of income).[9] Table 1 shows the actual position of the economies at four key dates. Countries that experienced a move of two or more income states between two consecutive dates are shown in bold (if moving to a higher income state) or italics (if moving to a lower income state).

Table 1 shows that a concentration on σ-convergence ignores some important historical issues. For example, despite the strong case made in the historical literature for the significance of the forces of convergence in the period 1870 to 1913, the empirical distributions suggest only Canada experienced a change of more than one income state between those two dates. Comparing 1950 to 1992, the (relative) losers were the UK and its two former southern dominions (Australia and New Zealand). Furthermore, the gains made by Germany and Austria appear to offset the losses they made between 1913 and 1950, whilst the one clear (relative) winner was Japan. Between 1870 and 1913 only six economies experienced a change in their income state and of these twice as many moved to a higher income state than moved to a lower income state. This contrasts with comparisons of 1913 to 1950 and of 1950 to 1992. In both of these cases nine economies experienced a change in their income state. Furthermore, economies that moved to a lower income state out-numbered those that moved to a higher income state (by a margin of two-to-one in the earlier period). Thus, there does not appear to be an obvious link between the

[9] The income ranges representing the five states are not imposed by the researchers but are derived on purely empirical grounds. First, all the annual standardised income observations are treated effectively as a single cross-section. The observations in this 'cross-section' are then ranked from the lowest to the highest observation and split into five equal states: each state contains the same number of observations. This gives us the values for the partition for each state. It also means that the size and values of the states is different for each sub-period, since the population is different for each sub-period. Alternative definitions of the income states are also possible and will be investigated in future research. We use Quah's TSRF (Time Series Random Field) which is an econometric shell that permits the calculation of transition probability matrices and ergodic distributions.

mobility shown in these empirical distributions and the σ-convergence shown in figure 1.

Table 2 shows the actual empirical distributions of the 24 economies in 1951, 1973 and 1992. There was slightly more movement between 1951 and 1973 than there was between 1973 and 1992 (respectively, sixteen and thirteen economies) and in both periods more economies moved to a higher income state than moved to a lower income state. That the degree of mobility shown in the two periods is fairly similar again contrasts to the significant difference between the periods shown in figure 1. Both tables 1 and 2 therefore suggest that the notion of convergence, and the forces that underlie it, is more complex than traditional stories tend to allow for.

3.3. Shapes

Another way of considering these snapshots of the empirical distributions is through the graphical representation of a kernel density estimator.[10] This has the advantage over a histogram representation in that it gives a smooth estimate. Figures 2 and 3 present kernel densities for the 17 and the 24 economies. Figure 2 shows that only for 1992 is there a very clear peak, and this was the period in which the variation in incomes was also smallest. Thus, 1992 most clearly tells a story of convergence. In Figure 3 we also see a move from a relatively flat distribution in 1950 to a more clearly peaked (if skewed) distribution in 1992.

Although the empirical distributions (and their associated kernel estimates) have some merit, they must be used with great care. For example, we would need to be sure that the years chosen for comparison were not atypical or affected by short term shocks. Furthermore, if we

---

[10] For an explanation of the kernel density estimator see the Technical Appendix. We have used

were trying to assess the impact of a particular regime on the distribution dynamics or trying to assess what the long run equilibrium associated with a particular regime would be, the approach taken above is far too limited and too crude. It tells us something only about relative income positions in a particular year but nothing about regimes' properties (or *inherent tendencies* that may or may not make for income convergence) as can be deduced from the analysis of the full set of annual income data. Herein lies the rationale for adopting distribution dynamics analysis. This is discussed below in terms of the mobility and the long term characteristics of the dynamic distribution.

4. Dynamic Distribution: Theory and Empirics

4.1. Three-Dimensional Representations: The Stochastic Kernel

One of the key issues we wish to address is: do relatively rich and poor countries remain relatively rich and poor countries over time? One way of measuring this would be to define a set number of income states (for example, rich, middle income and poor) and then to count the number of transitions out of one income state into another income state. This could then be formalised into a transition probability matrix. The matrix can then be used (either in its original form or in iterated versions) to discuss the degree of *mobility* and *persistence* within the distribution.[11] It could be argued, however, than the results may be sensitive to the (arbitrary) number of discrete income states chosen. Thus, below we consider stochastic kernels. The stochastic kernel (and its related contour plot) is a graphical representation of the transition probabilities which has the advantage that it does not rely on a fixed number of discrete states but

the Epanechnikov kernel.

[11] For an example of this approach see Epstein, Howlett and Schulze (1999). Figures 3 and 4 (pp.24-5) and pp.9-11 are of particular interest in terms of the following discussion as they outline two-dimensional representations of the dynamic distribution approach, emphasising the importance of changes in *shape* and *mobility*.

instead estimates a generalised form of the transition probability matrix in which renders the state space continuous.

A stochastic kernel can be generated for any length of transition. Figure 4 considers the stochastic kernel and contour plot for 5-year transitions in each of the three regimes for the 17 economies and figure 5 considers the same for the 24 economies in the two post-war periods.[12] Thus in each case, the income state of each economy in year t is compared to its income state in year t+5 and this is then averaged over the period of the regime.

How do we interpret the stochastic kernel? The stochastic kernel provides evidence both about the *shape* of and the *mobility* within the dynamic distribution. The horizontal axes (for *Period t* and *Period t + 5*) shows relative income, with 1.0 representing the standardised average level of income. Thus, a movement from left to right along either horizontal axis represents a move from the relatively poor to the relatively rich. The vertical axes measure the marginal probability density function. They effectively measure concentration or *clustering*. In terms of the shape, the key issue is whether or not the stochastic kernel has clear peaks or not. If there is a clear single peak this can be taken as evidence of convergence (and if the peak is centred on the 1.0 value of the *Period t + 5* horizontal axis this would show a convergence towards equality). If there is more than one peak this might be indicative of some form of club convergence. Furthermore, if this were also associated with a dip in the middle of the distribution this would suggest that *separation*, whereby the middle income economies move into either high or low income states, was an important underlying characteristic. Mobility can also be assessed by asking how the stochastic kernel lies relative to the 45-degree diagonal. If it is

[12] Any length of transition can be chosen, for example, in his 1997 paper Quah considered 15-year transitions. To be consistent across the different regimes being considered, we have chosen 5-year transitions. We also compared these to the 1-year transition stochastic kernels for each regime and found them to be robust (in the sense that there was no significant difference between the two different kernels).

concentrated along this diagonal then there is little sign of mobility: economies relative income in *Period t + 5* has not changed significantly since *Period t* (the relatively rich remain rich and the relatively poor remain poor).[13] A more interesting question is whether there is evidence of obvious shifts around the 45-degree diagonal. A counter-clockwise movement would represent the situation in which, relatively speaking, the rich were becoming poorer and the poor were becoming richer, thus indicating forces of convergence. At the extreme, this might take the form of over-taking with rich countries becoming poor and poor countries becoming rich. A clockwise movement would indicate the reverse: that the rich were becoming richer and the poor were becoming poorer, thus indicating that the forces of divergence were potentially more powerful. The contour plot of the stochastic kernel allows an easier identification of peaks and of movements of the distribution.[14]

4.2. Empirical evidence

What does figure 4 reveal about the distribution dynamics of the 17 advanced economies under the three different regimes? In terms of shape there is a distinction between the pre-1914 period and the later two periods in that the former is characterised by a single peak and the latter by twin peaks (although there is also evidence of clustering at the top end of the distribution in the former period). The single peak in the pre-1914 period is centred close to the 1.0 value of the *Period t + 5* horizontal axis which would tend to support the Williamson story of convergence. However the peak is also centred, more or less, on the 45-degree diagonal, which suggests that mobility was not a significant factor.

In the transwar period the lower income peak in this distribution is centred on the 45-degree diagonal whereas the upper income peak

---

[13] Obviously, with relatively short transition periods one would expect to find that most of the stochastic kernel would be concentrated along the 45-degree diagonal.

[14] Given the time horizon the kernel has been estimated for, we would not expect there to be

exhibits a clockwise movement. Moreover, mobility in this period is mainly concentrated in the tails of the distribution and exhibits counter-clockwise movement. Furthermore, despite the twin peaks the transwar stochastic kernel is flatter than that for the other two periods. Thus, overall the transwar period is one of conflicting signals (in which the forces of convergence and divergence jostle for attention), perhaps not surprisingly given the traumas of Maddison's 'bleak age'.

The twin peaks are clearer in the post-war period and are characterised by a counter-clockwise movement. Thus, the twin peaks are associated with some form of club convergence. The existence of twin peaks raises significant questions about the findings of Mankiw, Romer and Weil (1992) that there was convergence within the OECD. However, it is not inconsistent with the more recent findings of Bernard and Durlauf (1995) or Crafts and Mills (2000), which caution against stories of strong convergence.[15]

The story told by figure 5 about the group of 24 economies is also interesting. The Golden Age shows evidence of twin peaks but this characteristic is even stronger in the Post-Golden Age. The dip in the middle of the distribution in the latter period also suggests that this is partly the result of separation. Indeed, the contour plot for the post-Golden Age period is indicative of what Durlauf and Johnson (1995) called 'basins of attraction'. The other important difference is that whereas in the Golden Age period the distribution is shifting counter-clockwise relative to the 45-degree line, in the post-Golden Age period it is shifting clockwise. Thus, in the earlier period there is some evidence of convergence whereas the latter period is more reminiscent of polarisation, whereby the relatively rich get richer and the poor become poorer.

---

major rotations.

[15] Crafts and Mills, however, are unwilling to accept 'the idea of convergence clubs, in which (different) Augmented-Solow specifications apply' but argue that their 'rejection of strict convergence and of common trends in all countries can be explained by differences in technology and, linked to this, "social capability"' (pp.85-7).

4.3. Long Run Equilibrium Distributions

If we believe that each of the sub-periods identified here do indeed represent different regimes it would also be useful to know what the long run equilibrium of each regime was. Although the stochastic kernel uses information on the actual empirical observations to generate its graphical representation of the distribution dynamics it is less useful in this context. To find the steady state solutions of a regime we need to use a discrete analysis of the transition probabilities. It is well known that an important feature of any transition probability matrix is that it will yield a unique long run equilibrium condition, the *ergodic distribution* (see the Technical Appendix). This is a central concept in the analysis since it permits gauging the 'convergence' properties of historical regimes. It allows us to estimate the extent to which particular regimes were conducive to convergence processes. We estimated one-year transition probability matrices for each of the data sets and periods under consideration using five income states (see discussion above) and also calculated their ergodic distributions.[16] The ergodic distribution suggests what the shape of the long run equilibrium distribution would look like.

Table 3 shows the ergodic distributions for the 17 and 24 economies associated with each period. We shall first consider the evidence for the 17 advanced economies. According to this evidence, the long run equilibrium of the pre-1914 regime was not one characterised by convergence. The long run equilibrium position shows two distinct plateaux, which suggests that stratification was at least as strong a force as persistence.[17] The situation is even more bland for the transwar

[16] As stated previously this was done using TSRF. A detailed historical example of the calculation and use of transition probability matrices (and ergodic distributions) is provided by Epstein, Howlett and Schulze (1999). The operation of the TSRF programme, as it pertains to this material, is further treated in Technical Appendix B of that paper.

[17] It should be noted that persistence was a prevalent feature in the one-year transition probability matrix for this period.

regime: the ergodic distribution is effectively flat across the first four income states and then tails off in income state 5. However, it should be noted that the shape of this distribution, which points neither to convergence nor divergence, cannot be interpreted as merely an outcome of persistence. Rather, it seems to have been associated with a considerable degree of intra-distributional 'churning'.[18] This is consistent with the history and historiography of this period. It is a period that experienced three of the most significant shocks in the twentieth century (the two World Wars and the Great Depression). Furthermore, the impact of these shocks, and the economic reaction to each shock, differed across individual economies. Thus, it is not surprising that we can not detect a clear regime in this period.[19] Finally, the post-war regime offers an example of convergence in that there is a clear peak in the middle of the distribution and the lowest points are at the extremes of the distribution.

Turning to the 24 economies, the long run equilibrium of the Golden Age regime shows a clear peak in the middle of the distribution (income sates 3 and 4 account for almost two-thirds of the distribution). In contrast, the long run equilibrium of the Post-Golden Age regime shows increasing density positively associated with a movement up the income state scale. This is an example of a peak that does not indicate convergence (indeed, the stochastic kernel indicated that it was a period of a form of club convergence in which polarisation, or the forces of divergence, was an important factor).

[18] The sum of the off-diagonal values in the transition matrix for the transwar period is the highest amongst the three sub-periods, indicating a high degree of the economies' mobility between income states. Whilst there is a lot of upward and downward movement within the distribution and there are rank order changes, these do not translate into a long run equilibrium distribution displaying pronounced uni- or multi-modality.

[19] In an unpublished paper (available from the authors on request), we attempted to abstract from the war shocks and considered only the period 1922-38. The ergodic distribution for this regime exhibited clear signs of uni-modality (that is, convergence). This suggests that the World Wars, rather than the Great Depression, are the important shocks in terms of the distribution dynamics of the regime.

5. Conclusions

In discussing the results we need to distinguish between the empirically observed **period dynamics**, and in particular what this reveals about mobility, and **the long-run equilibrium of the regime** as represented by the ergodic distribution, for which shape is the key characteristic. For the 17 advanced OECD economies a distribution dynamics approach suggests that the pre-1914, transwar and post-war periods were each distinctive and different. The empirical stochastic kernel for the pre-1914 period was effectively uni-modal but it did not exhibit significant mobility.[20] This was further reinforced by the ergodic distribution that suggested that the long run equilibrium of this regime was not one of convergence but rather was characterised by the forces of persistence and stratification. This makes it difficult to reconcile with the Williamson story of strong convergence in this period caused by globalisation. In terms of mobility, the transwar period provided no strong evidence for either convergence or divergence, or rather the evidence could be interpreted to provide some support for both, and the dominant force would appear to have been churning. This is again consistent with the relatively flat shape of the ergodic distribution. In the post-war period, the empirical stochastic kernel suggests that mobility was driven by the forces of convergence. The evidence from this kernel of some form of club convergence is perhaps surprising given the selection bias involved. However, the ergodic distribution suggests the post-war regime is characterised by a uni-modal steady state. Finally, the analysis of the distribution dynamics of the 24 economies in the post-war period potentially poses even more fundamental questions. It suggests that convergence was a feature of the Golden Age but that it was a temporary phenomenon (or one specific to that regime). Once problems were

[20] The one-year transition probability matrices that show that the pre-1914 period was characterised by a comparatively low degree of mobility within the international income distribution also confirm this result.

encountered the regime gave way to one in which separation, divergence and polarisation came to the fore: in the Golden Age the relatively rich got poorer and the relatively poor got richer but in the Post-Golden Age the poor gathered around one peak and the rich around another one. If this is then related to the empirical distributions of table 2 it begs the question as to how far the convergence of the Golden Age was an artefact of post-war reconstruction and catch-up.

Technical Appendix

A stochastic process is a system that evolves over time according to probabilistic laws.[21] Such systems are defined by one or more 'states', or groupings of the data. The dynamics are described by the *transition probabilities* that individual elements of the process will move from one state to another from period t, say, to period $t + n$, $n = 1, 2, \ldots$ Quah has applied this model to the dynamics of the distribution of income levels in the world economy. His technique permits a 'law of motion' for the stochastic process to be deduced, which in turn permits the long-run behaviour of the system to be described.

Systems defined by discrete states are known as *Markov chains*.[22] In the distribution dynamics literature five states are commonly employed, each containing a proportion of the economies in the distribution. If the vector $\boldsymbol{x_t}$ gives the proportion of economies in each state in year t, then the dynamics can be modelled by a first order matrix autoregression in probabilities, $\boldsymbol{x_t} = \boldsymbol{M_t}\,\boldsymbol{x_{t-1}}$. $\boldsymbol{x_t}$ is then interpreted as a probability vector. Each element of the matrix M_t gives the probability of an economy moving from state r to state s from period t to period $t + 1$. The elements of $\boldsymbol{M_t}$ are the *transition probabilities*; $\boldsymbol{M_t}$ is known as the *transition probability matrix*, or transition matrix, each of whose rows sums to unity. The leading diagonal of the matrix shows the *persistence* of the system, i.e., the probability that an element in state r in period t will remain in that state in period t+1. Off-diagonal elements show the system's *mobility*: that is, the tendency to change state and the associated transition probabilities.

The transition matrix can be used to describe the evolution of the whole system over time: that is, if $\boldsymbol{x_{t+1}} = \boldsymbol{M}\,\boldsymbol{x_t}$ (dropping the subscript on M), then by iteration, $\boldsymbol{x_{t+q}} = \boldsymbol{M}^{q}\,\boldsymbol{x_t}$, q ³ 1. $\boldsymbol{M}^{q}$ gives the transition probabilities

[21] Formal expositions of the mathematics underlying stochastic processes are given in Chung (1960), Doob (1953) and Stokey and Lucas (with Prescott) (1989).

[22] An intuitive presentation of the Markov chain and the decomposition of the transition matrix is given in Cox and Miller (1990), Section 3.2. For applications of this model, see Proudman,

from the current period q periods ahead. Because $\boldsymbol{M}$ is a transition probability matrix its largest eigenvalue is unique. Consequently $\boldsymbol{M}^{q}$ converges to a matrix of rank one, say $\boldsymbol{M}$¢ *,* whose rows are all identical. Mathematically, there is a *spectral representation,* or decomposition, of $\boldsymbol{M}$, yielding the linear combination of matrices

$$\boldsymbol{M} = a_1 \boldsymbol{M}¢ + a_2 \boldsymbol{M}¢¢ + ¼$$

The a_j are functions of the eigenvalues of $\boldsymbol{M}$, so for an n-state system $\boldsymbol{M}$ is a linear combination of the ergodic matrix (i.e., row vector) $\boldsymbol{M}$¢ and n-1 transient matrices. Because $a_1 = 1$ and $\boldsymbol{M}$¢ is a row vector, it is constant on iteration; because the $a_j < 1$ for $j > 1$, the transient components go to zero. That is,

$$\boldsymbol{M}^{q} = \boldsymbol{M}¢ + \mathbf{0} + \mathbf{0} + \ldots$$

Following Quah's notation, $\boldsymbol{M}$¢ satisfies the equation $\boldsymbol{x}$¥ = $\boldsymbol{M}$¢ $\boldsymbol{x}$¥, where $\boldsymbol{x}$¥ is the ergodic vector. $\boldsymbol{x}$¥ is interpreted as the *limiting distribution*, or steady state of the process; that is, the limit of the above difference equation as $q \to \infty$.

Example: Consider a two–state Markov chain. The transition probabilities from time t to time t+1 can be denoted

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \end{array}.$$

In a hypothetical case for the Maddison data set state 0 can be interpreted as being 'poor', state 1 as 'rich'. Suppose that the program finds the following transition probabilities:

$$\mathbf{P} = \begin{array}{c} \\ \textit{poor} \\ \textit{rich} \end{array}\begin{array}{c} \begin{array}{cc} \textit{poor} & \textit{rich} \end{array} \\ \begin{bmatrix} 0.668 & 0.332 \\ 0.231 & 0.769 \end{bmatrix} \end{array}.$$

The leading diagonal, *diag*(0.668, 0.769), shows that the hypothetical

---

Redding and Bianchi (1997).

system is largely persistent. To find the state at time *t+k*, given the state at time *t,* it is sufficient to raise the matrix **P** to the power of *k*: that is , $\mathbf{P}\times\mathbf{P}\times\mathbf{P}\times\ldots$, *k* times. In the above case, for example,

$$\mathbf{P}^2 = \begin{bmatrix} 0.668 & 0.332 \\ 0.231 & 0.769 \end{bmatrix}\begin{bmatrix} 0.668 & 0.332 \\ 0.231 & 0.769 \end{bmatrix} = \begin{bmatrix} 0.523 & 0.477 \\ 0.332 & 0.668 \end{bmatrix}$$

(to three decimal places). Raising **P** to further powers gives, for example,

$$\mathbf{P}^5 = \begin{bmatrix} 0.420 & 0.580 \\ 0.404 & 0.596 \end{bmatrix}$$

and

$$\mathbf{P}^{10} = \begin{bmatrix} 0.410 & 0.590 \\ 0.410 & 0.590 \end{bmatrix}$$

which illustrates the convergence property of Markov chains. $\mathbf{P}^{10}$ is a matrix of rank 1, i.e., a row vector. It is interpreted as the steady state of the system represented by the transition probabilities in **P**; the 'ergodic vector' [0.410, 0.590] (to three significant figures).

Denote the eigenvalues of **P** as l_1 and l_2. Then by a result of Belman (1960) there is a 'diagonal' or 'spectral' representation

$$\mathbf{P} = \mathbf{Q}\begin{bmatrix} l_1 & 0 \\ 0 & l_2 \end{bmatrix}\mathbf{Q}^{-1}.$$

Additionally,

$$\mathbf{P}^n = \mathbf{Q}\begin{bmatrix} l_1{}^n & 0 \\ 0 & l_2{}^n \end{bmatrix}\mathbf{Q}^{-1},$$

Now, l_1, l_2 are the solutions of the characteristic equation of **P**, $|\mathbf{P} - \lambda\mathbf{I}| = \mathbf{0}$, i.e., $(1-a-l)(1-b-l)-ab=0$, giving $l_1 = 1, l_2 = 1-a-b$. $l_1 \neq l_2$ if $a+b \neq 0$; $l_2 < 1$ if $a+b \neq 0$ or $a+b \neq 2$.

We can find

$$\mathbf{Q} = \begin{bmatrix} 1 & a \\ 1 & b \end{bmatrix} \quad \Rightarrow \quad \mathbf{Q}^{-1} = \frac{1}{a+b}\begin{bmatrix} b & a \\ 1 & -1 \end{bmatrix}$$

such that

$$\mathbf{P}^n = \frac{1}{a+b}\begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b}\begin{bmatrix} a & -a \\ -b & b \end{bmatrix}$$

Clearly the first term is constant; the second goes to zero as *n* increases. In the hypothetical example, a = 0.332, b = 0.231, so

$$\mathbf{P}^n = \begin{bmatrix} 0.410 & 0.590 \\ 0.410 & 0.590 \end{bmatrix} + \frac{(0.437)^n}{0.563}\begin{bmatrix} 0.332 & -0.332 \\ -0.231 & 0.231 \end{bmatrix}.$$

Hence, since $(0.437)^{10} = 2.54 \times 10^{-4}$, or zero to three significant figures, $\mathbf{P}^{10}$ is as above. This argument can be generalised to processes with an arbitrary number of states. Since the largest eigenvalue of any probability matrix is always unity, a steady state is guaranteed.

Kernel Estimates

The disadvantage of the discrete, *n*–state Markov model is that the number of states is chosen arbitrarily. The kernel–density estimator is a generalisation of the discrete model in which the *n* tends to infinity, rendering the state space continuous. The histogram is a discrete analogue of the kernel estimate of single densities. Data are divided into disjoint class intervals with the bar centred at the midpoint of the interval. The height of each bar reflects the number of observations in the interval. As an extension of the histogram, the kernel density estimate permits the class intervals to overlap. In effect the interval, which in this case is called a 'window', slides along the range of the observations, and centre point estimates are made, the width of the window being known as 'bandwidth'. Unlike the histogram, the kernel weights each observation by its distance from the centre point. The weighted observations are summed to give the height of the ordinate at each point at which the kernel is being estimated. Estimates are smoother, and therefore more easily comparable with known parametric distributions.

The transition probability matrix can be seen as a histogram of joint densities. The *stochastic kernel* is a generalisation of this.[23] It can be represented either as a two dimensional contour plot or as an orthogonal projection onto a surface in three dimensions. Each point of the surface is interpreted as a probability. That is, the stochastic kernel estimate is simply the continuous analogue of the transition probability matrix.

Formally, following Quah (1997) Section 4, let m and n be probabilities, and let *A* be a window; then the stochastic kernel is the mapping $M_{(m,n)}$ which satisfies

$$m\ (A) = \int M_{(m,n)}\,(y, A)\, dn\,(y)$$

subject to certain restrictions. That is, for a given *A,* the count *M(y,A)* is weighted by *dn(y)* and summed over all possible values of *y*, giving the fraction of economies ending up in state *A* regardless of their initial state. The restrictions, which are discussed in Quah (1997), guarantee that $m\,(A)$ is a Lebesgue integral, that is, a weighted sum of data points in the window *A*. The stochastic kernel gives a complete description of transition probabilities from state *y* to any other state.[24]

[23] Applications of the stochastic kernel are discussed in Quah(1997) and Durlauf and Quah (1998).

[24] Further related readings include Doob (1953), Bellman (1960), Chung (1960), Stokey and Lucas (1998).

Journal of Applied Econometrics, 10, pp.97-108.

Broadberry, S. N. (1996). 'Convergence: What the historical record shows', in B. van Ark and N.F. R. Crafts (eds.), *Quantitative aspects of post-war European growth* (Cambridge), pp.327-346.

Chung, K.L. (1960). *Markov Chains with Stationary Transition Probabilities* (Berlin).

Cox, D.R. and Miller, H.D. (1990). *The Theory of Stochastic Processes* (London).

Doob, J.L. (1953). *Stochastic Processes* (New York).

Durlauf, Steven N and Johnson, Paul A. (1995). 'Multiple regimes and cross-country growth behaviour', *Journal of Applied Econometrics*, vol. 10, pp. 365-84.

Durlauf, S. and Quah, D. (1998). 'The New Empirics of Economic Growth'. *Centre for Economic Performance Discussion Paper No. 384*. London School of Economics and Political Science.

Epstein, P., Howlett, P. and Schulze, M-S. (1999). 'Income distribution and convergence: the European experience, 1870-1992', *LSE Working Papers in Economic History* 52/99.

Good, D.F., and Ma, T. (1998). 'New estimates of income levels in Central and Eastern Europe, 1870-1910', in F. Baltzarek et al. (eds.), *Von der Theorie zur Wirtschaftspolitik – ein österreichischer Weg* (Stuttgart), pp.147-168.

Japan Statistical Association. (1998). *Historical Statistics of Japan*, vol. 2 (Tokyo).

Maddison, A. (1995). *Monitoring the world economy, 1820-1990* (Paris).

Mankiw, N.G., Romer, D., and Weil, D. (1992). 'A contribution to the *Quarterly Journal of Economics*, 107, pp.407-38.

Mills, T.C. and Crafts, N.F.R. (2000). 'After the Golden Age: a long run perspective on growth rates that speeded up, slowed down and still differ', *Manchester School*, Vol. 68, pp.68-91.

Ohkawa, K., Shinohara, M., with L. Meissner. (1979). *Patterns of Japanese economic development. A quantitative appraisal* (New Haven, London).

Prados de la Escosura, L. (1995). 'Spain's Gross Domestic Product, 1850-1993: Quantitative conjectures. Appendix', Universidad Carlos III, Working Paper 96-06, Economics Series 02.

Proudman, J., Redding, S. and Bianchi, M. (1997). 'Is International Openness Associated with Faster Economic Growth?' *Bank of England working paper series No. 63*. Bank of England.

Quah, D. T. (1993). 'Empirical cross-section dynamics in economic growth *European Economic Review* 37, pp.426-34.

Quah, D. T. (1996). 'Twin peaks: growth and convergence in models of *Economic Journal* 102, pp.1045-55.

Quah, D. T. (1997). 'Empirics for growth and distribution: stratification, polarization, and convergence clubs', *Journal of Economic Growth* 2, pp.27-59.

Romer, P. (1986). 'Increasing returns and long run growth', *Journal of Political Economy* 94, pp.1002-37.

Schulze, M-S. (2000, forthcoming). 'Patterns of growth and stagnation in the 19th Century Habsburg economy', *European Review of Economic History*.

Economic History Review 51, pp.252-67.

Tortella, G. (1994). 'Patterns of economic retardation and recovery in south-western Europe in the nineteenth and twentieth centuries', *Economic History Review* 47, pp.1-21.

Williamson, J.G. (1996). 'Globalization, Convergence, and History', *Journal of Economic History* 56, pp.277-306.

Table 1. Empirical distribution by income state, 17 OECD economies, 1870 - 1992

Income state			**1870**			
1	FIN	JAP				
2	CAN	ITA	NOR	SWE	SPA	
3	AUT	DEN	FRA	GER		
4	BEL	NED	USA			
5	AUS	NZ	UK			
			1913			
1	FIN	JAP	SPA			
2	ITA	NOR	SWE			
3	AUT	FRA				
4	BEL	**CAN**^{2}	DEN	GER	NED	UK
5	AUS	NZ	USA			
			1950			
1	*AUT*^{2}	ITA	JAP	SPA		
2	FIN	*GER*^{2}				
3	BEL	FRA	NOR	NED		
4	AUS	CAN	DEN	**SWE**^{2}	UK	
5	NZ	USA				
			1992			
1	FIN	*NZ*^{4}	SPA			
2	*AUS*^{2}	ITA	*UK*^{2}			
3	**AUT**^{2}	BEL	FRA	NED	NOR	SWE
4	CAN	DEN				
5	**GER**^{3}	**JAP**^{4}	USA			

Bold indicates a move to a higher income state compared to the previous year shown; *italic* shows a move to a lower income state. The subscripts show by how many income states the economy has moved.

Table 2. Empirical distribution by income state, 24 economies, 1951 - 1992

Income state				1951				
1	GRE	HUN	JAP	POR	SPA	TUR		
2	AUT	CZE	GER	IRE	ITA	FIN		
3	BEL	FRA	NOR					
4	NED							
5	AUS	CAN	DEN	NZ	SWE	SWZ	UK	USA
				1973				
1	HUN	TUR						
2	CZE	GRE	IRE	NOR	POR	SPA		
3	AUT	BEL	FIN	ITA	**JAP**^{2}	*NZ*^{2}	*UK*^{2}	
4	AUS	CAN	DEN	FRA	**GER**^{2}	NED	SWE	
5	SWZ	USA						
				1992				
1	CZE	GRE	HUN	TUR				
2	FIN	IRE	NZ	POR	SPA			
3	AUS	ITA	UK					
4	AUT	BEL	FRA	**NOR**^{2}	SWE			
5	CAN	DEN	GER	**JAP**^{2}	NED	USA	SWZ	

Bold indicates a move to a higher income state compared to the previous year shown; *italic* shows a move to a lower income state. The subscripts show by how many income states the economy has moved.

Figure 1. Coefficients of variation of real GDP per capita, 1870-1992

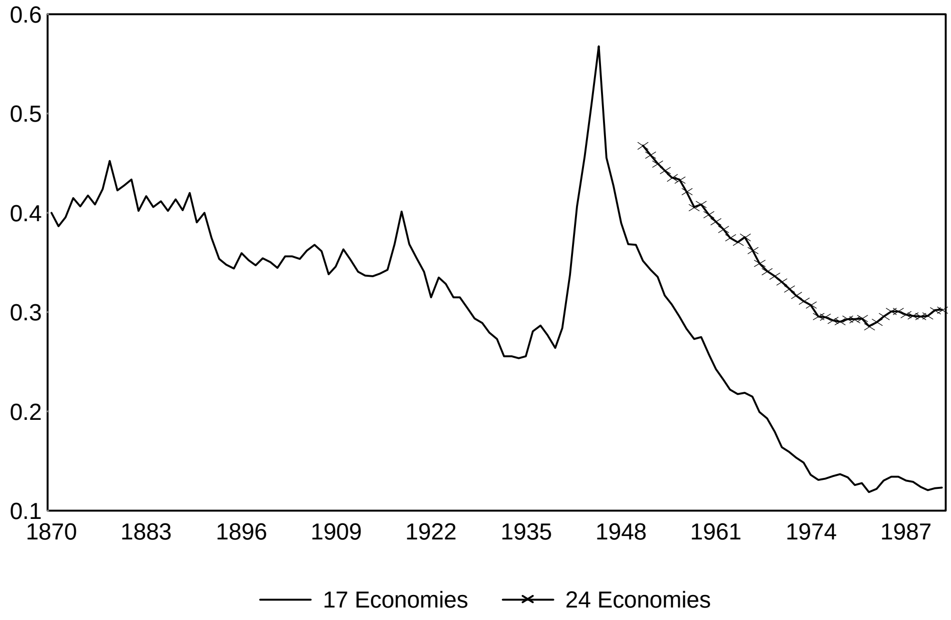


Figure 2. Densities of real GDP per capita in 17 OECD economies, 1870-1992

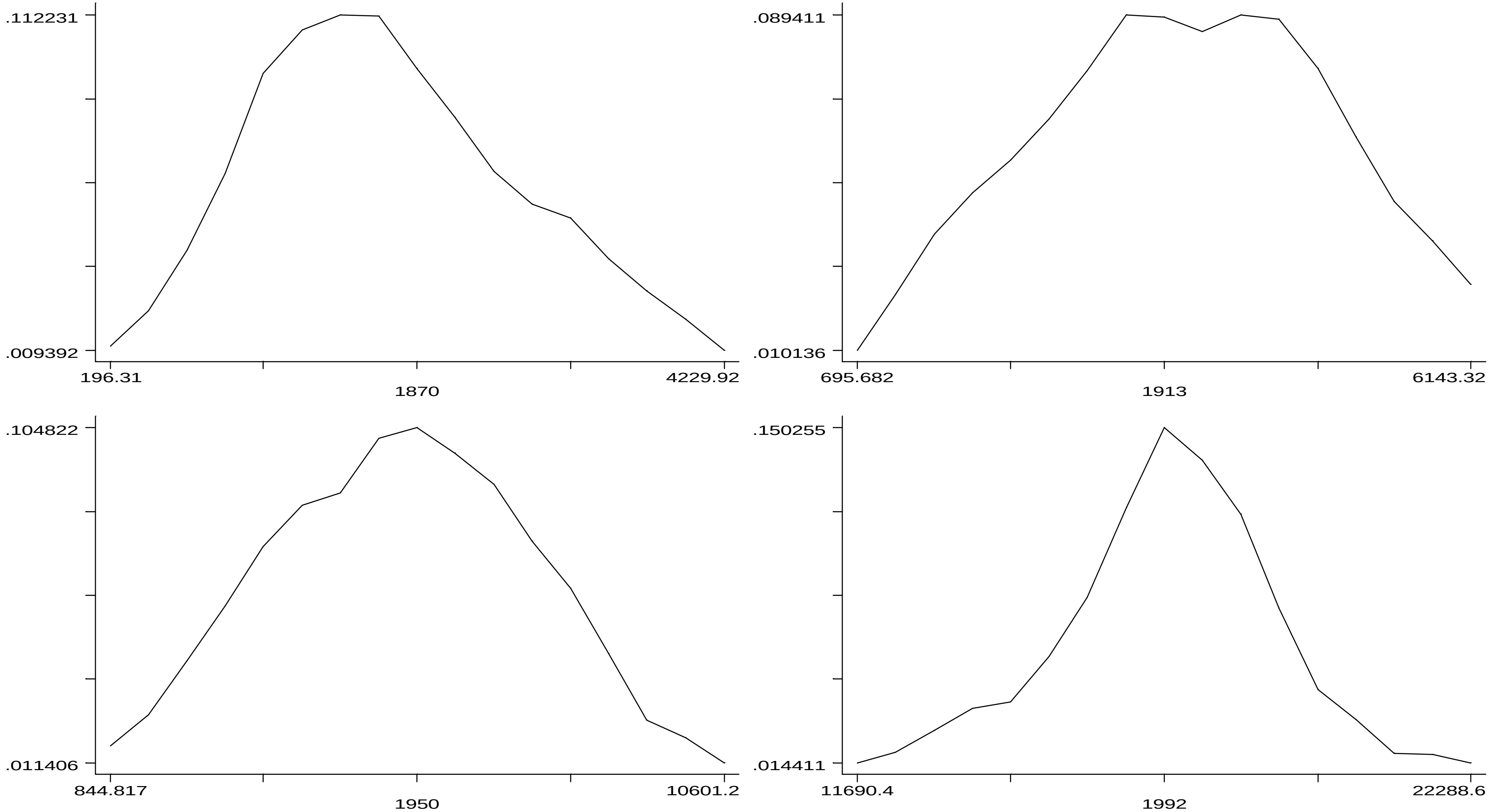


Figure 3. Densities of real GDP per capita in 24 economies, 1951 and 1992

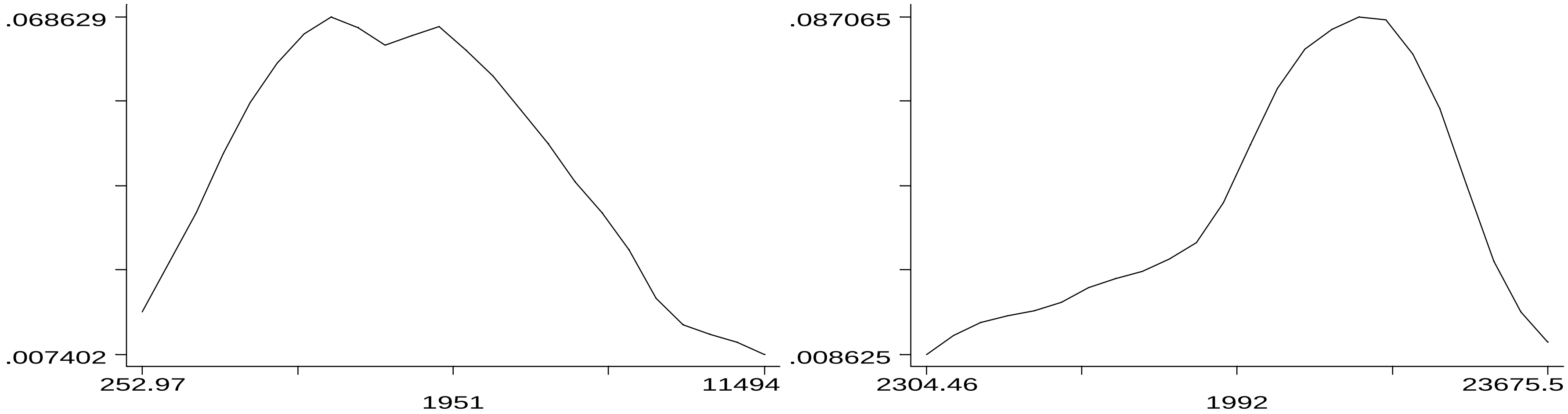
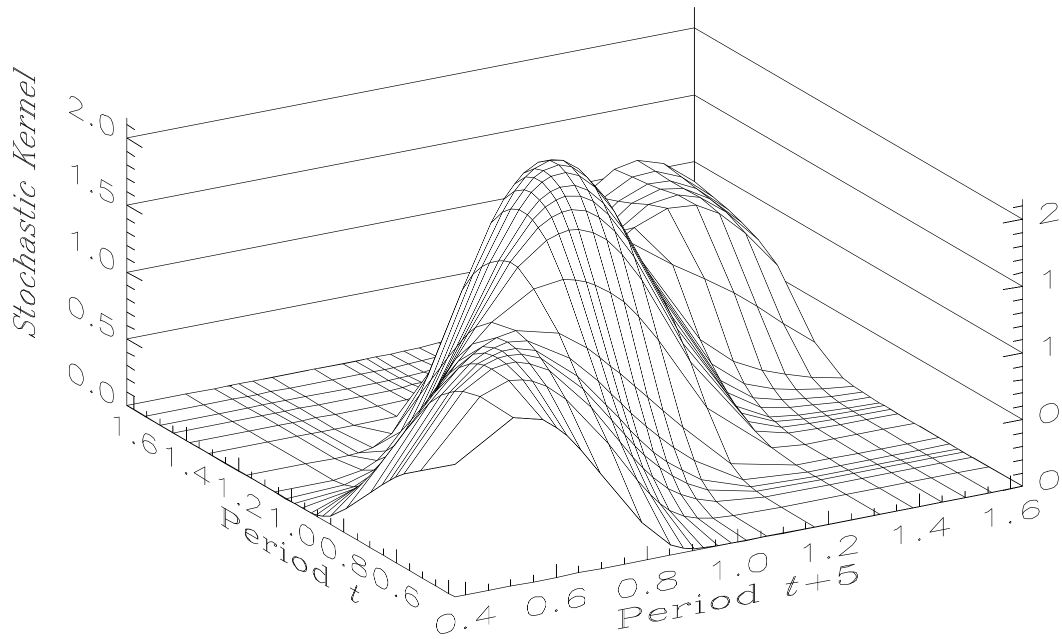


Figure 4a. 17 Advanced Economies, 1870-1913



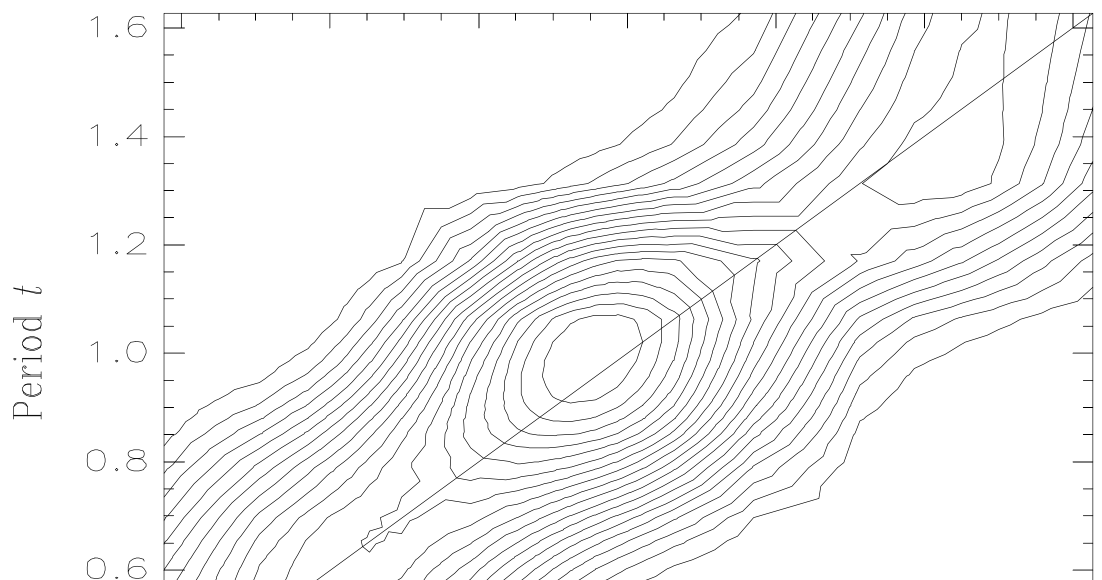
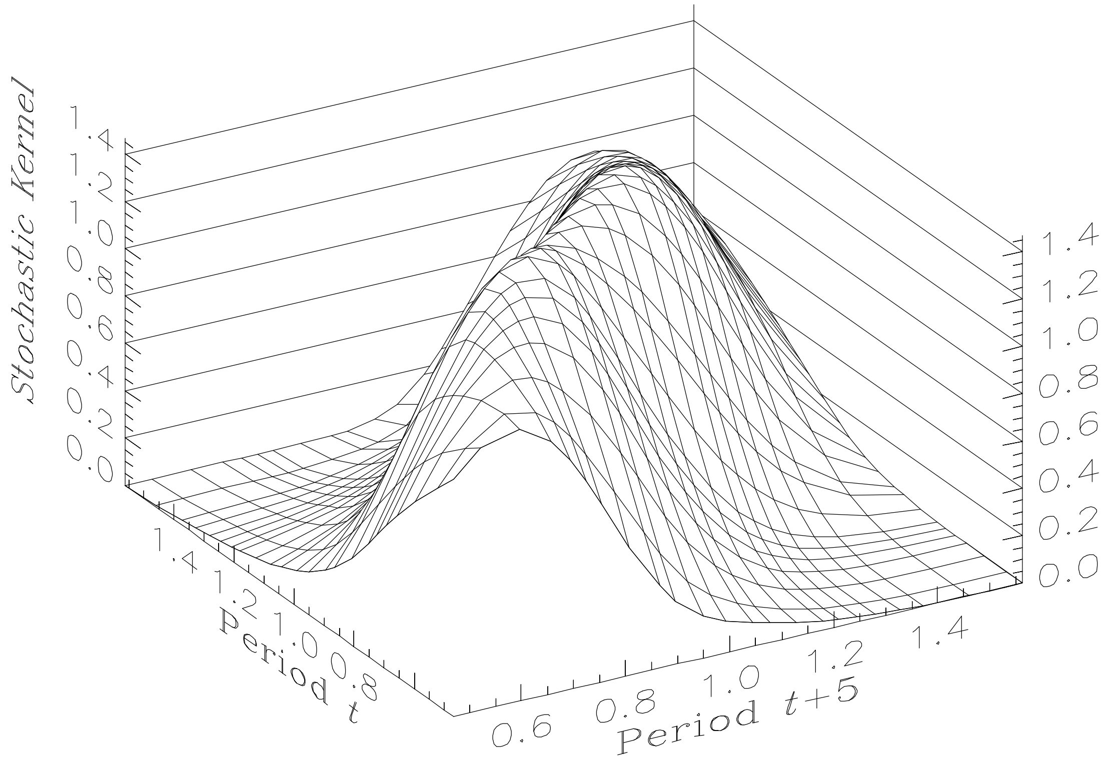


Figure 4b. 17 Advanced Economies, 1914-1950



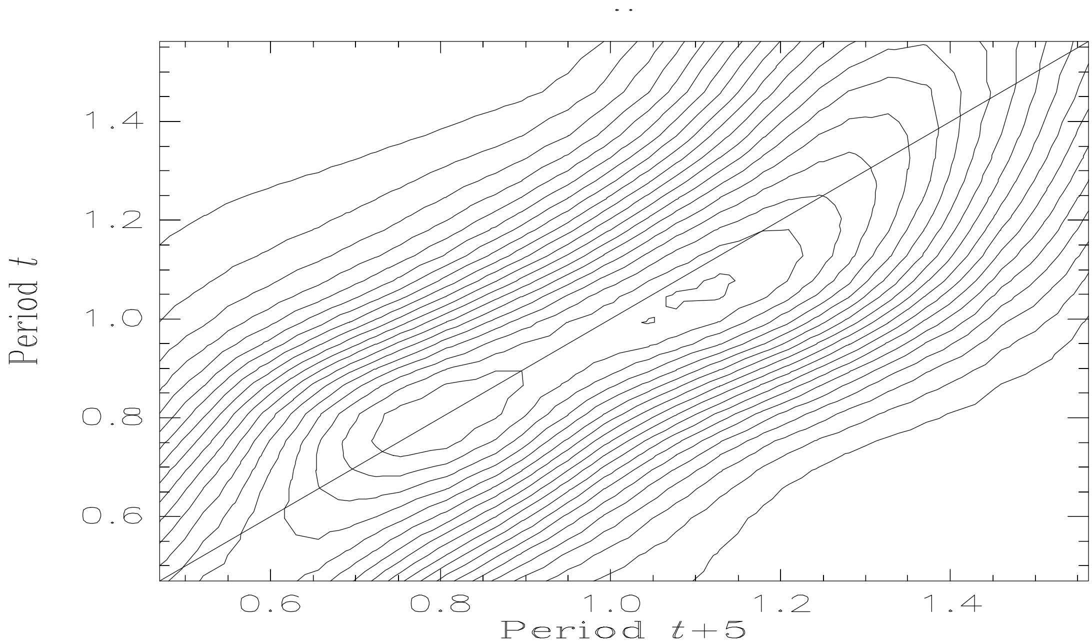


Figure 4c. 17 Advanced Economies, 1951-1992

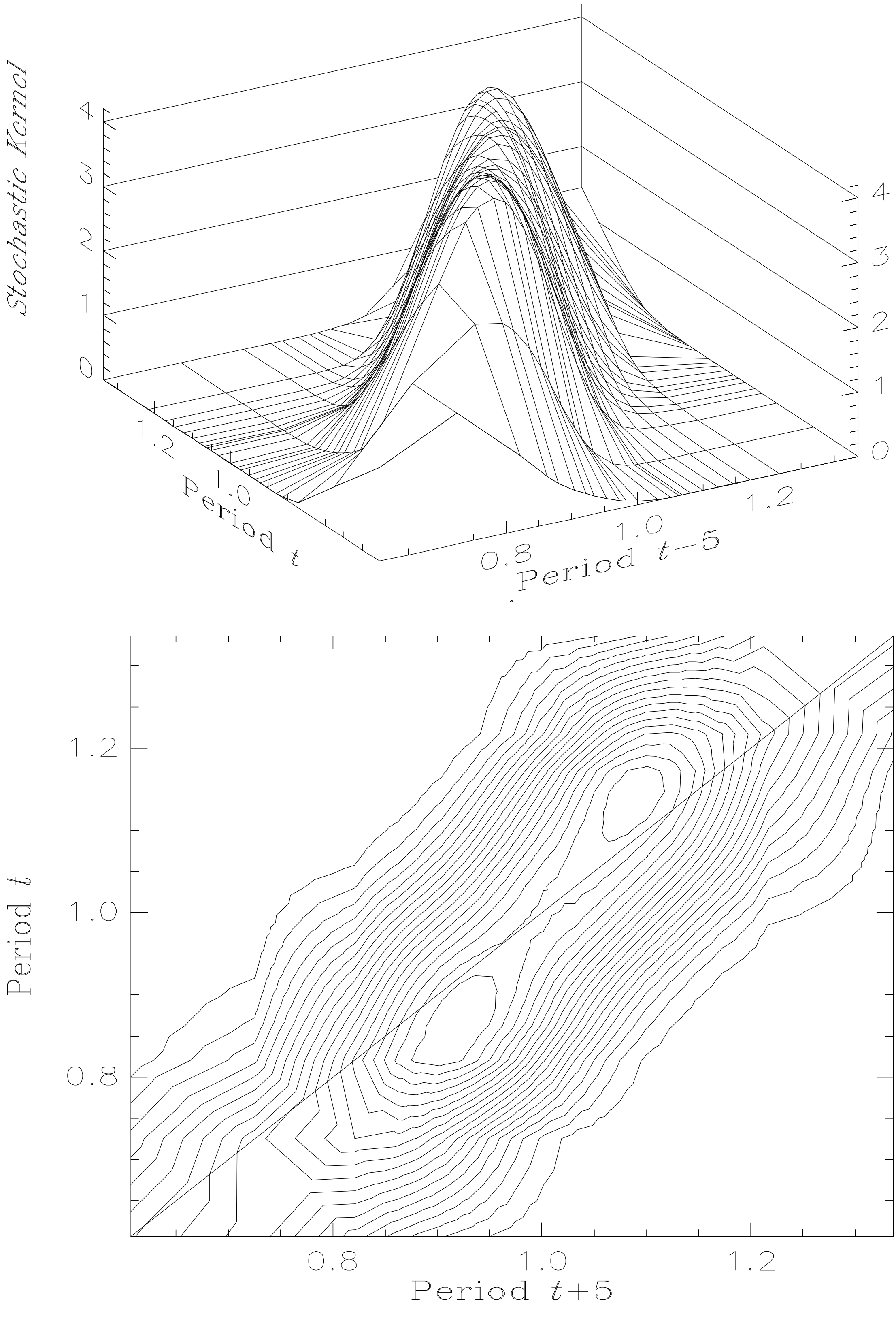
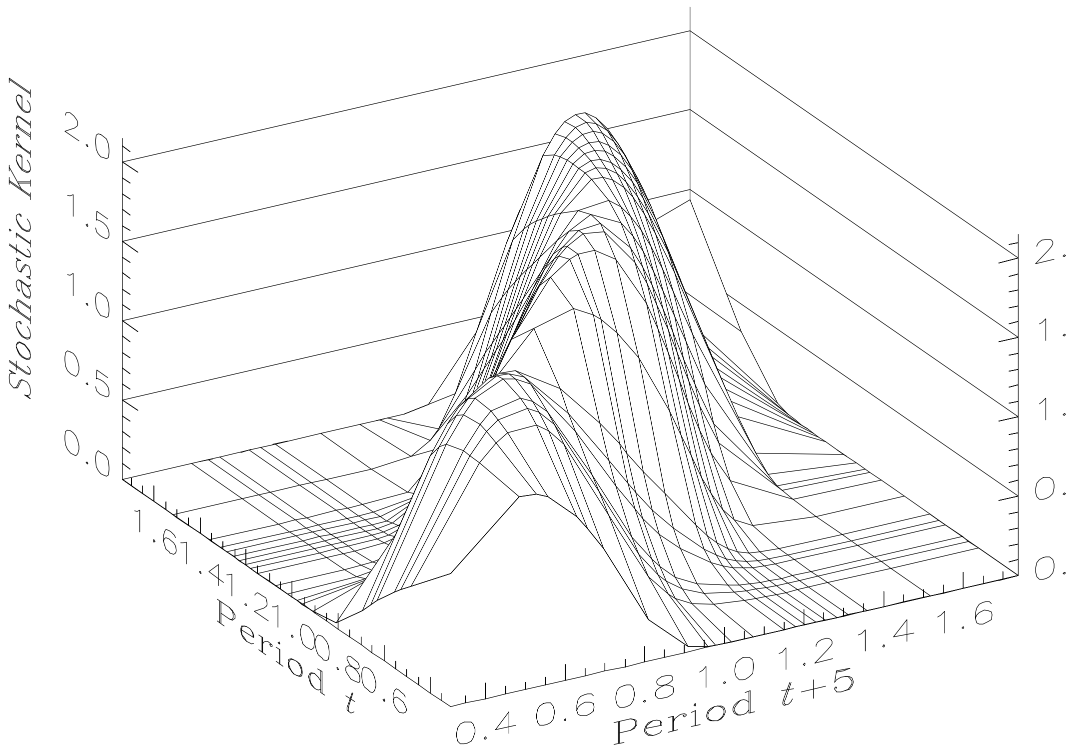


Figure 5a. 24 Economies, 1951-1973



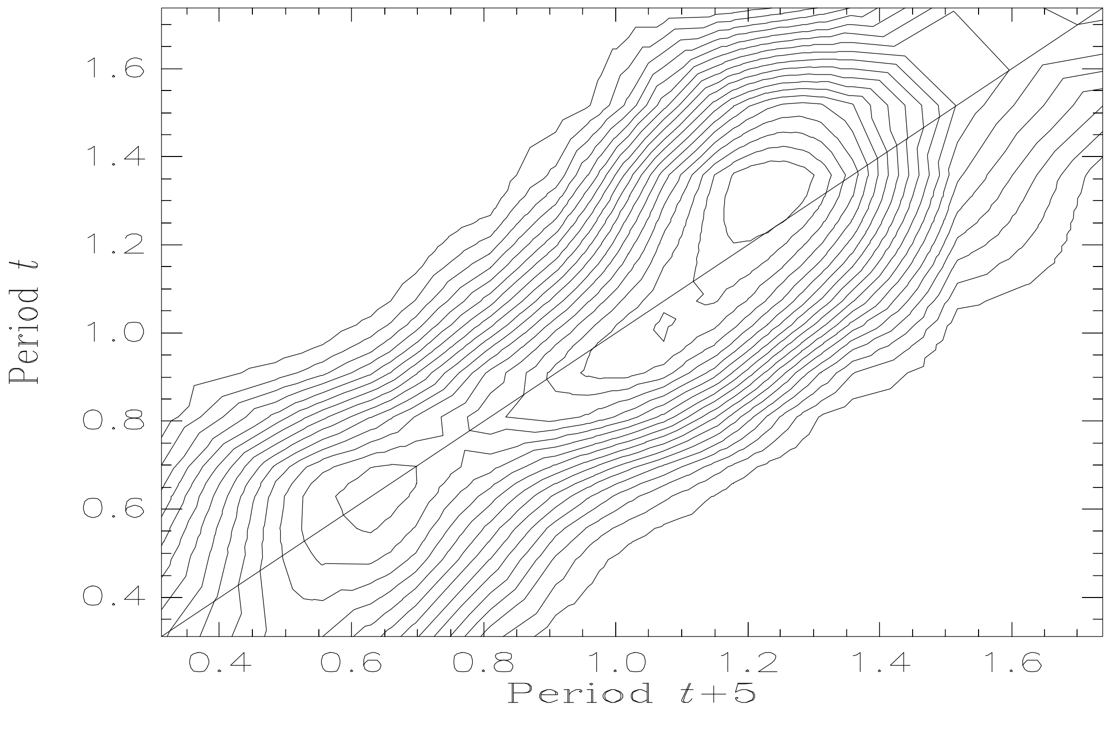
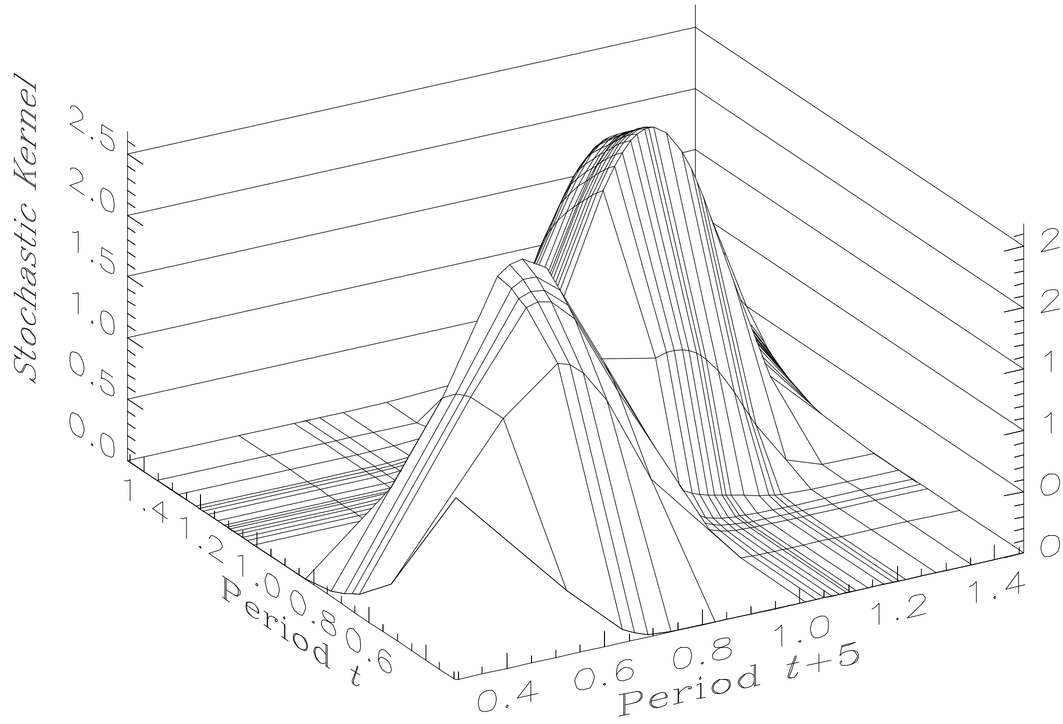


Figure 5b. 24 Economies, 1974-1992



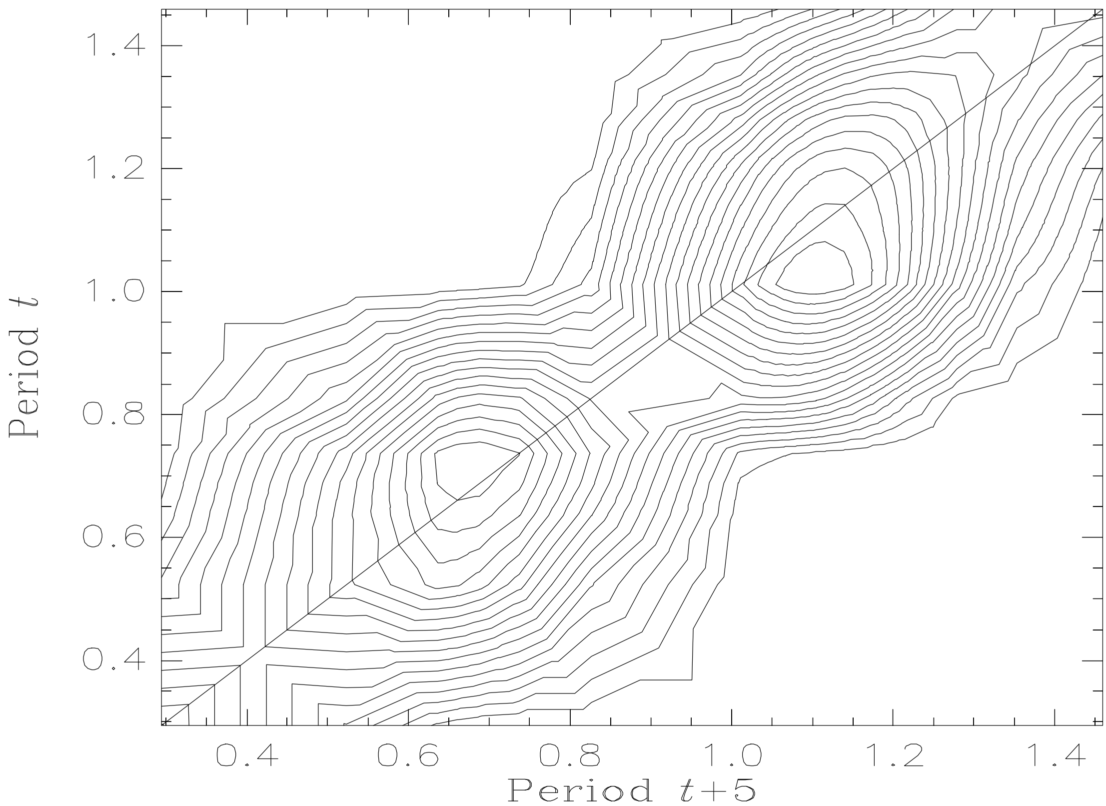


Table 3. Ergodic Distributions

	Income States				
	1	**2**	**3**	**4**	**5**
17 advanced economies, 1870 - 1992					
1870 – 1913	.146	.131	.144	.280	.298
1914 – 1950	.218	.205	.202	.201	.174
1951 – 1992	.156	.203	.259	.228	.153
24 advanced economies, 1951 - 1992					
1951 – 1973	.038	.182	.313	.319	.148
1974 – 1992	.103	.153	.212	.241	.291