Review of Income and Wealth
Series 56, Number 1, March 2010

MIXTURE MODELS, CONVERGENCE CLUBS, AND POLARIZATION

by Maria Grazia Pittau* and Roberto Zelli

Sapienza University of Rome

and

Paul A. Johnson

Department of Economics, Vassar College

We argue that modeling the cross-country distribution of per capita income as a mixture distribution provides a natural framework for the detection of convergence clubs. The framework yields tests for the number of component distributions that are likely to be more informative than "bump hunting" tests and includes a method of assessing the cross-component immobility necessary to imply a correspondence between components and convergence clubs. Applying this approach to Penn World Data for the period 1960 to 2000 we find evidence of three component densities. We find little cross-component mobility and so interpret the multiple mixture components as representing convergence clubs. We document a pronounced tendency for the strength of the bonds between countries and clubs to increase and show that the well-known "hollowing out" of the middle of the distribution is largely attributable to the increased concentration of the rich countries around their component means.

1. Introduction

There has been a great deal of interest in the shape and evolution of the cross-country distribution of per capita income in recent years. Much of this interest arises from the relationship between those characteristics of the distribution and the neoclassical convergence hypothesis. That hypothesis states that initial conditions have no implications for long-run outcomes so that all countries will converge to a common level of GDP per capita regardless of where they begin.[1] An alternative hypothesis is that initial conditions do matter in the long run and that countries with similar initial conditions exhibit similar long-run outcomes, so forming "convergence clubs"—groups of countries that converge locally but not globally. One possible manifestation of the presence of convergence clubs is multiple modes in the cross-country distribution of per capita income, with each mode corresponding to a convergence club. Multimodality is, however, not enough to

Note: We thank, without implicating, Gordon Anderson, Steven Durlauf, Daniel Henderson, Ann Owen, and participants at the Israel Science Foundation Research Workshop on "Income Polarization: Measurement, Determinants, Implications" for comments on earlier versions of this paper. Also we would like to thank the editors and two anonymous referees for their useful comments and suggestions. Paul Johnson thanks the Lucy Maynard Salmon Research Fund at Vassar College for financial support and DSPSA, Sapienza Università di Roma, for their hospitality, during preparation of this paper. Maria Grazia Pittau and Roberto Zelli gratefully acknowledge financial support from Sapienza research funds.

*Correspondence to: Maria Grazia Pittau, DSPSA, Sapienza University of Rome, P.le Aldo Moro 5, 00185 Roma, Italy (grazia.pittau@uniroma1.it).

[1]Durlauf *et al.* (2005) provide a survey of the many guises taken by the convergence hypothesis and of their myriad empirical implementations.

imply the existence of convergence clubs. That requires immobility within the distribution so that countries in the vicinity of a mode tend not to move to that of another mode.

Most of research investigating the shape of the cross-country distribution of per capita income has employed kernel estimation methods. See, for example, Quah (1996, 1997), Bianchi (1997), Jones (1997), Henderson *et al.* (2008), and others. Bianchi (1997) and Henderson *et al.* (2008) present various tests of the hypothesis of a unimodal distribution against that of a multimodal distribution. They are able to reject the null in most cases. Both papers also find little mobility between the modes they identify. Together, these findings support the existence of convergence clubs.[2] Applications of mixture models, a semi-parametric alternative to the kernel approach, have been less numerous. The only application to the cross-country distribution of which we are aware is Paap and van Dijk (1998), although Tsionas (2000) uses mixture models to study the distribution of per capita output across the U.S. states, while Pittau (2005) and Pittau and Zelli (2006) use them to study the distribution of per capita incomes across EU regions.[3] The findings of Paap and van Dijk are consistent with those of Bianchi (1997) and Henderson *et al.* (2008).

The mixture approach has important advantages over the kernel approach in the current application. Mixture models express the density of a random variable as the weighted average of a finite number of component densities with specified functional form. The parameters to be estimated are the number of, the weights attached to, and the parameters of, the component densities. When used to describe the cross-country distribution of per capita income, the components in a mixture model can be interpreted as corresponding to the basins of attraction in the dynamic process describing the evolution of per capita income. Multiple components then, like multiple modes, can be indicative of multiple basins of attraction.

However, just as there is no reason why multiple components in a mixture distribution will manifest as multiple modes, there is no reason why multiple basins of attraction will manifest as multiple modes. The mixture approach is able to detect the presence of multiple components in a distribution even if that multiplicity does not manifest itself as multimodality. As multimodality is not necessary for the existence of convergence clubs, used as part of a test of the convergence hypothesis, the mixture approach can thus provide a test with more power to detect convergence clubs than the kernel approach, which relies on the detection of multiple modes for rejection of the convergence hypothesis.

As with the kernel approach, the interpretation of multiple components as indicative of convergence clubs also requires an analysis of the mobility within the distribution, which in this case means that between the components. Again, this

[2]While these two papers do find evidence of multiple modes, the general perception that there are two modes in the cross-country distribution of per capita income owes its existence to the many authors who have contributed to the so-called "twin peaks" literature.

[3]Since completing the initial draft of this paper, we have become aware of a paper by Holzmann *et al.* (2007) which employs a similar approach to that employed here. Our statistical method is somewhat more general since it allows the component variances not to be equal. We find that these variances differ both across components and over time and have an important interpretative role.

can be accomplished naturally within the mixture model framework, providing another improvement on the relatively *ad hoc* methods of mobility analysis employed in the kernel based studies. The estimated mixture model parameters enable computation of the conditional probabilities that each entity belongs to each component. These probabilities can be used to assign entities to components as well as to gauge the strength of the affinity between the entity and the components. The propensity of entities to change their assigned components over time provides a measure of within-distribution mobility.

In this paper we use finite mixture models to investigate the number of components in the cross-country distribution of per capita income over the 1960 to 2000 period. In addition to the improvements over the studies based on kernel estimation mentioned above, the primary contribution is the improvement in the methodology over that of Paap and van Dijk (1998) who choose the number of components to be two *a priori* based on the bimodality of histograms of their data. This procedure may not detect all components as components do not imply modes. Indeed, we find strong evidence of three rather than two components. The next section of the paper outlines our analytical framework and describes the data we use. Section 3 presents our results and the final section offers our conclusions.

2. Analytical Framework and Data

2.1. *Mixture Models*

Mixture models provide an appealing semi-parametric structure in which to model unknown distributional shapes. The *m*-component mixture model specifies the density of a random variable as

$$f(x, m, \Theta_m, \Pi_m) = \sum_{j=1}^{m} \pi_j f_j(x, \boldsymbol{\theta}_j), \tag{1}$$

where $f_j(x, \boldsymbol{\theta}_j)$ is a probability density function with parameter vector $\boldsymbol{\theta}_j$, for $j = 1, \ldots m$, $\Theta_m = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_m)$, the π_j are the mixing proportions with $\pi_j > 0$ for $j = 1, \ldots, m$, $\sum_{j=1}^{m} \pi_j = 1$, and $\boldsymbol{\Pi}_m = (\pi_1, \pi_2, \ldots \pi_m)$.

Given m and the functional forms of the component densities, $f_j(x,\boldsymbol{\theta}_j)$, the parameters of the model can be estimated by the method of maximum likelihood. We do so using an iterative fitting by maximum likelihood (ML) via the expectation–maximization (EM) algorithm (Dempster *et al.*, 1977). Each iteration comprises an expectation step (E-step) followed by a maximization step (M-step). The EM algorithm seems to be superior to the other procedures in finding a local maximum of the likelihood function (McLachlan and Peel, 2000). We make the assumption that the component densities are normal with unequal variances so that $f_j(x, \boldsymbol{\theta}_j) = N(x; \mu_j, \sigma_j^2)$, the normal density function with mean μ_j and variance σ_j^2, for $j = 1, \ldots, m$. This is not, however, as restrictive as it may seem because any continuous density can be well approximated by a mixture of normal densities (Marron and Wand, 1992). Moreover, the normal distribution is especially easy to interpret in this application as μ_j is the mean per capita income in component j and σ_j^2 measures the within-component variation in per capita incomes.

We take two approaches to the selection of m, the number of components. The first follows McLachlan (1987) and is a likelihood ratio test (LRT) of the null hypothesis $m = m^*$ against the alternative $m = m^* + 1$. For this test, the distribution of the LRT statistic under the null hypothesis is estimated by bootstrap methods as the conditions necessary for the LRT statistic to have the usual asymptotic χ^2 distribution do not hold.[4]

For each m^*, B bootstrap samples are drawn from the mixture distribution $f(x, m^*, \hat{\Theta}^*_m, \hat{\Pi}^*_m)$ where the parameter values are those estimated using the original sample. An m^* component and an $m^* + 1$ component mixture model are estimated for each sample by the method of maximum likelihood, and the usual LRT statistic is computed. The significance level of the sample LRT statistic is then computed as $1-\frac{r}{B+1}$ where r is number of replications with an LRT statistic less than the sample LRT statistic.

The second approach to selecting the number of components considers the goodness of fit of the estimated mixture model by comparison of a kernel estimate of the density of the data and its expected value under the null hypothesis that the population density is a mixture of m^* normal distributions. The alternative hypothesis is that the true number of components exceeds m^*. The test statistic is a measure of agreement between these two densities, the estimated integrated squared error (ISE) statistic:

$$\hat{J} = \int_x \left[\hat{f}(x) - \hat{E}\hat{f}(x)\right]^2 dx. \tag{2}$$

The estimated mean $\hat{E}\hat{f}(x)$ is a convolution of a kernel with a mixture of normal components $\hat{E}\hat{f}(x) = K_h * \hat{f}(x) = \frac{1}{h}\int_u K\left(\frac{x-u}{h}\right) f\left(u; \hat{\Psi}_0\right) du$ When the kernel function is Gaussian, the convolution collapses into a mixture of normal densities with means equal to μ_j and variances equal to $\sigma_j^2 + h^2$. Therefore the statistic results as:

$$\hat{J} = \int_x \left[\hat{f}(x) - \sum_{j=1}^{m^*} \hat{\pi}_j N\left(x; \hat{\mu}_j, \hat{\sigma}_j^2 + h^2\right)\right]^2 dx, \tag{3}$$

where h is the bandwidth used to compute $\hat{f}(x)$, the kernel estimate of $f(x)$, the true density of x. We select h using the Sheather and Jones (1991) method, which is widely recommended due to its overall good performance (Jones *et al.*, 1996). While asymptotic results for the distribution of $\hat{J}$ are available, we follow Fan (1995) because of our small sample size, and we estimate the distribution using a parametric bootstrap procedure in which the bootstrap samples are drawn from the mixture distribution $f(x, m^*, \hat{\Theta}^*_m, \hat{\Pi}^*_m)$ where the parameter values are those estimated using the original sample. The significance level of the sample $\hat{J}$ is

[4]Other approaches, like the modified LRT, derive a relatively simple asymptotic null distribution of the likelihood ratio test. See, distinctively, Ghosh and Sen (1985), and more recently Chen *et al.* (2004) and Chen and Kalbfleisch (2005). However, the implementation of such modified LRT does not alter the findings of this section.

computed as $1-\frac{r}{B+1}$ where B is the number of bootstrap replications and r is the number of replications with $\hat{J}$'s less than the sample $\hat{J}$.

We apply both the LRT and ISE tests sequentially, beginning with the null hypothesis $m^* = 1$, continuing to that of $m^* = 2$ if the $m^* = 1$ null is rejected, and so on. In both cases, we set m equal to the smallest m^* for which we are unable to reject the null hypothesis $m = m^*$. Once m is chosen, the parameter vectors $\boldsymbol{\Theta}_m$ and $\boldsymbol{\Pi}_m$ can be estimated, enabling us to study the properties of the m component densities. The π_j can be interpreted as the unconditional probability that X_i, observation i, is a draw from component j. The conditional probability of that event is given by:

$$\zeta_{ji} = \frac{\pi_j f_j(X_i, \boldsymbol{\theta}_j)}{\sum_{j=1}^{m} \pi_j f_j(X_i, \boldsymbol{\theta}_j)}. \tag{4}$$

These probabilities can be used to assign observations to components by assigning observation i to that component with the largest estimated ζ_{ji}, computed using equation (4) with the π_j and the $\boldsymbol{\theta}_j$ replaced by their estimates. Given a panel of data, mobility can be studied by noting the propensity of the assignment of entity i to change over time.

2.2. *Mixtures and Club Convergence*

While we do not attempt to estimate a specific economic model in this paper, many authors have described models that could generate a cross-country distribution of per capita output that is well explained by an m-component finite mixture model.[5]

Generically, these models imply a law of motion for per capita output, x_t, such as $x_{t+1} = \phi(x_t, \varepsilon_t)$ as shown in the top panel of Figure 1 (adapted from figure 2 in Galor, 1996) for the case when ε_t, a random disturbance, equals zero. Fixing ε_t at zero yields a non-stochastic dynamic system with two locally-stable steady states at x_L and x_H and an unstable steady state at x_M.

Countries starting with $x_t < x_M$ would converge to x_L and those starting out with $x_t > x_M$ would converge to x_H so that initial conditions (or "history") would serve to determine long-run outcomes. Allowing ε_t to vary in the appropriate way could produce a cross-country distribution of per capita output with density $f(x)$ as illustrated in the second panel of Figure 1. The distribution is bimodal with each mode corresponding to one of the stable steady states in the non-stochastic case. Initial conditions will again play a role determining long-run outcomes, but in this case they select the distribution of per capita output. Depending on the distribution of ε_t, countries that begin in the vicinity of the left-hand mode will tend to remain there, while counties that begin in the vicinity of the right-hand mode will

[5]Azariadis (1996) and Galor (1996) provide comprehensive surveys of theoretical mechanisms capable of producing this convergence club behavior, highlighting those such as: different saving propensities out of labor and capital income; low elasticities of substitution between capital and labor; demographic transitions to sharply lower fertility rates as wages rise; external increasing returns; and external effects from social interactions, to list a few. See also Bloom *et al.* (2003).

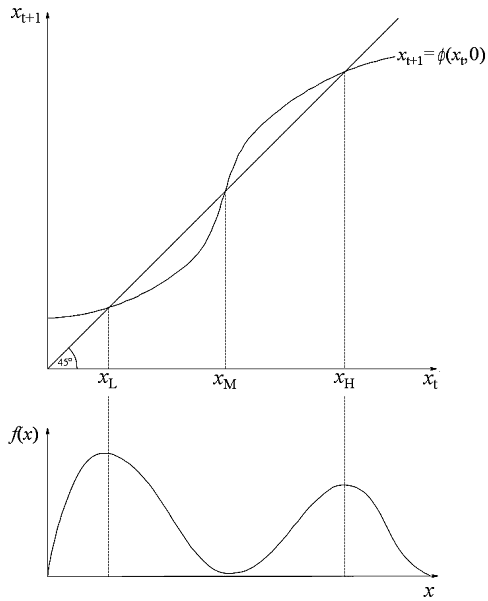

Figure 1. Multiple Steady States and Convergence Clubs (adapted from figure 2 in Galor, 1996)

tend to remain there. There will be little mobility within the distribution, but occasionally countries will transit from one basin of attraction to the other if they are subject to realizations of ε_t that are particularly (un)favorable.[6] As mentioned in the introduction, the distribution need not be (as sharply) bimodal as that shown and the ability of the mixture approach to detect multiple components even when they do not manifest as distinct modes is a part of our motivation for using the approach in this case.

2.3. *Data*

The per capita income data used is real GDP per worker (RGDPWOK) from the Penn World Table (PWT) Version 6.1 (Heston *et al.*, 2002). As Durlauf *et al.* (2005) argue, GDP per worker accords more closely than GDP per capita with the dependent variable of interest in most growth models.[7] The sample consists of data on the 102 countries—all of those for which data is available for the entire 1960 to 1995 period. Following Durlauf *et al.* (2005) we exclude the middle-eastern oil producing countries and Luxemburg. For the year 2000 data we use data from 1998 for 98 of the 102 countries, with that for the remaining four being extrapolated from the 1997 data.[8] The alternative, using the countries for which actual 2000 data is available, would reduce our dataset to 89 countries. We estimate a

[6]The dynamics in the stochastic case could be described by a stochastic kernel as discussed in, for example, Quah (1996).

[7]The number of workers ". . . is usually a census definition based of economically active population" (Data Appendix to PWT 6.1, dated 10/18/02, p. 11).

[8]As the data for each year are analyzed independently, any errors caused by this extrapolation will be confined to the 2000 data.

mixture model for each of the nine years 1960, 1965, . . . , 2000. The variable used in our analysis is RGDPWOK relative to its workforce-weighted average over the 102 countries in the sample. Using the PWT 6.1 mnemonics, the workforce for each country, in each year, was computed as POP*RGDPCH/RGDPWOK.

3. Results

3.1. *Number of Components*

Table 1 reports the LRT statistics and the corresponding bootstrapped *p*-values for testing the null hypothesis of $m = m^*$ components for m^* ranging from 1 to 4. In each year the value of the LRT statistic implies rejection, at conventional significance levels, of the null hypotheses $m = 1$ and $m = 2$ but not that of $m = 3$. That is, the null of $m = 3$ is selected over the alternative that $m = 4$. Moreover there is no tendency for the selected number of components to fall over time as would be suggested by a tendency for the LRT statistics for the $m = 2$ null hypothesis to fall. To the contrary, if there is any tendency at all for the selected number of components to change, it is for them to rise as evinced by the rise in the LRT statistics for the $m = 3$ null hypothesis, although we are never able to reject this hypothesis.

Similarly, Table 2 presents the results of the statistical testing procedure using the goodness of fit test (ISE) based on the Sheather and Jones smoothing parameter. These results are robust to the selection of the smoothing parameter h used in the kernel estimation of the unknown density function. In fact we do not reject the hypothesis of $m = 3$ components for a wide range of values of h that also includes the well known Silverman's rule of thumb bandwidth and the bandwidth obtained by the least squares cross-validation method Silverman (1996). These results are also entirely consistent with those from the LRT procedure, lending support to the conclusion that a mixture of three normal densities offers the preferred description of the cross-country distribution of output per worker.

The finding of three (or, more generally, more than one) mixture components is not enough to imply the existence of multiple convergence clubs in the cross-country distribution of per capita income. That requires an additional analysis of

TABLE 1

The Choice of the Number of Components According to the Likelihood Ratio Test

	$m^* = 1$		$m^* = 2$		$m^* = 3$		$m^* = 4$	
Year	LRT	p-value	LRT	p-value	LRT	p-value	LRT	p-value
1960	64.18	0.000	24.21	0.042	3.80	0.736	0.01	1.000
1965	56.14	0.000	34.91	0.026	3.33	0.804	0.00	0.998
1970	59.93	0.000	28.20	0.036	5.35	0.574	1.00	0.978
1975	51.89	0.000	37.52	0.022	3.72	0.642	0.01	1.000
1980	62.53	0.000	20.87	0.048	0.51	0.978	0.00	1.000
1985	47.06	0.002	35.08	0.028	2.17	0.932	0.00	1.000
1990	55.18	0.000	45.28	0.024	9.12	0.206	3.03	0.942
1995	64.47	0.000	45.17	0.020	10.17	0.192	0.58	0.992
2000	61.74	0.000	46.15	0.016	11.22	0.154	2.91	0.978

TABLE 2

THE CHOICE OF THE NUMBER OF COMPONENTS ACCORDING TO THE GOODNESS OF FIT TEST

	$m^* = 1$		$m^* = 2$		$m^* = 3$		$m^* = 4$	
Year	$\hat{J}$	p-value	$\hat{J}$	p-value	$\hat{J}$	p-value	$\hat{J}$	p-value
1960	10.76	0.000	2.11	0.000	0.21	0.736	0.04	0.960
1965	10.80	0.000	3.34	0.000	0.22	0.776	0.09	0.812
1970	10.61	0.000	1.19	0.006	0.38	0.365	0.11	0.713
1975	10.22	0.000	3.10	0.000	0.23	0.642	0.04	0.850
1980	9.28	0.000	2.23	0.000	0.06	0.954	0.05	0.849
1985	9.09	0.000	2.83	0.001	0.42	0.156	0.33	0.057
1990	11.58	0.000	3.55	0.000	0.50	0.192	0.17	0.579
1995	11.32	0.000	3.62	0.000	0.47	0.219	0.13	0.678
2000	11.51	0.000	2.42	0.000	0.48	0.112	0.14	0.673

Note: The estimated ISE, $\hat{J}$, is multiplied by 100.

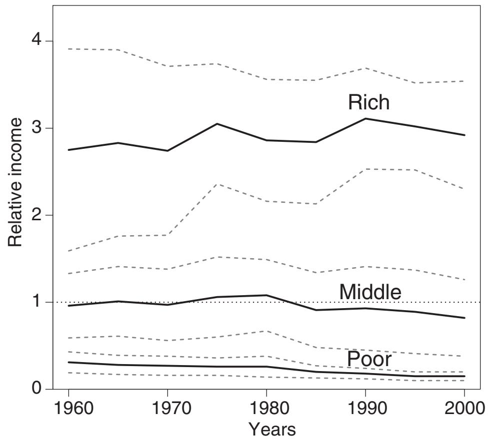

Figure 2. Groups' Means Over Time

the mobility of the basins of attraction, which we undertake in Section 3.3 after a discussion of the evolution of the components over the sample period.

3.2. *Evolution of the Distribution and its Components*

The estimation of the previous section produces estimates of the mean μ_j and variance σ_j^2 of per capita GDP for each of the three components which we label "poor," "middle," and "rich" according to the estimated means with $\hat{\mu}_{\text{poor}} < \hat{\mu}_{\text{middle}} < \hat{\mu}_{\text{rich}}$.

Figure 2 plots the means against time (the solid lines) along with dashed lines that indicate the intervals containing 80 percent of the probability mass of each component. That is, the dashed lines are $\hat{\mu}_{\text{poor}} \pm 1.282 \times \hat{\sigma}_{\text{poor}}$, $\hat{\mu}_{\text{middle}} \pm 1.282 \times \hat{\sigma}_{\text{middle}}$, and $\hat{\sigma}_{\text{rich}} \pm 1.282 \times \hat{\sigma}_{\text{rich}}$, where $\hat{\sigma}_j$ is the estimated standard deviation of component j for j = "poor," "middle," "rich."

As Figure 2 shows, over the sample period the mean of the poor component, $\hat{\mu}_{\text{poor}}$, fell steadily so that, in 2000, it was about half of its 1960 value, although because the 1960 value is so low—about 30 percent of the sample mean—this fall is small in absolute terms. The estimated means of the middle and rich components are slightly more volatile than that of the poor component, with $\hat{\mu}_{\text{rich}}$ having an upward trend over the sample period while $\hat{\mu}_{\text{middle}}$ finishes the sample period slightly below where it began. The gap between the rich and poor components, measured as $\hat{\mu}_{\text{rich}} - \hat{\mu}_{\text{poor}}$, increases by about 14 percent over the sample period while $\hat{\mu}_{\text{rich}} - \hat{\mu}_{\text{middle}}$ increases by about 17 percent.

There are also important changes in the dispersion of the countries around the component means, especially in the case of the rich component. Over the sample period the estimated standard deviation for this component, $\hat{\sigma}_{\text{rich}}$, falls by almost 50 percent, with about half of the fall occurring between 1970 and 1975 and a further quarter occurring between 1985 and 1990. This is shown in Figure 2 by the narrowing of the interval containing 80 percent of the mass of this component to 60 percent of its 1960 value in 1975 and subsequently to 50 percent of its 1960 value in 1990.

This phenomenon, and the relative stability of the estimated standard deviation of the middle component, which rises by about 30 percent over the sample period, combine to open a region of low probability mass between the middle and rich components. This is evident in the successive panels of Figure 3 as the deepening of the antimode at a value of relative output per worker of about two. This figure shows the estimated kernel and mixture densities for GDP per worker in each year as well as the constituents of the estimated mixture distribution, i.e. the $\hat{\pi}_j f_j(x, \hat{\boldsymbol{\theta}}_j)$ for $j =$ "poor," "middle," "rich."[9] As panels (a), (b), (c), and (d) in Figure 3 show, the antimode is evident in 1960 and it remains substantially unchanged until 1975 when it becomes much deeper. Panels (e), (f), and (g) show that after 1975 the antimode was again substantially unchanged through 1980 and 1985 until it again become much deeper in 1990.

While the gap between the rich component and the middle component means does rise—by about 17 percent—over the sample period, the dominant cause of the observed "hollowing out" of the middle of the cross-country distribution of output per worker seems to be the decrease in the within-component variation in the rich component. As this decrease could reflect, in part at least, compositional changes, we have more to say about it in Section 3.4 after we discuss mobility across the components.

The variance of the poor component falls by almost 60 percent. As Figure 3 shows, the net effect of this and the smaller rise in the variance of the middle component is the appearance in 1965 of an antimode at a value of relative output per worker of about two-thirds. This antimode persists at various depths throughout the remainder of the sample period but is never very deep compared to the mode immediately to its right (at a value of relative output per worker slightly

[9]These are the same densities used to compute the $\hat{J}$ statistics discussed above. The $\hat{\pi}_j f_j(x, \hat{\boldsymbol{\theta}}_j)$ are not individually labeled due to space considerations but there ought not be any resultant ambiguity as $\hat{\pi}_{\text{middle}} f_{\text{middle}}(x, \hat{\boldsymbol{\theta}}_{\text{middle}})$ lies always to the right of $\hat{\pi}_{\text{poor}} f_{\text{poor}}(x, \hat{\boldsymbol{\theta}}_{\text{poor}})$ and so on.

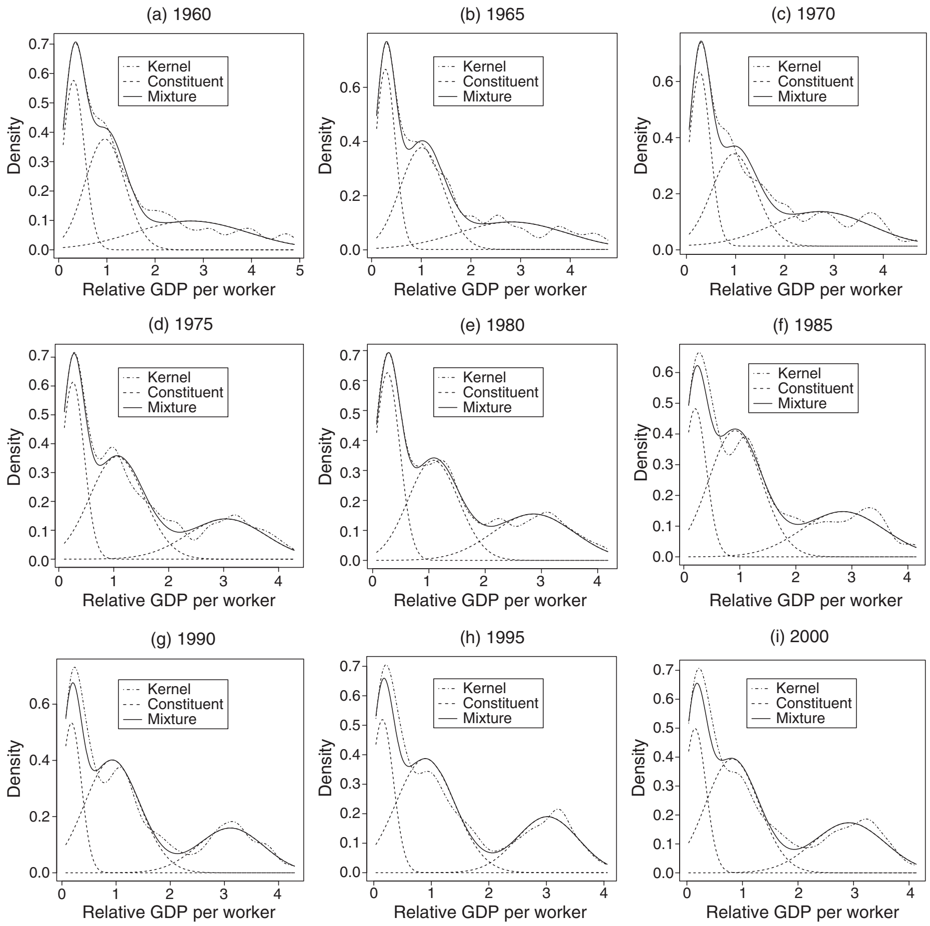

Figure 3. Kernel Density Estimation and the Bias-Adjusted Three-Components Mixture Model Fit Over the Period 1960–2000

above unity). The importance of this phenomenon in the evolution of the distribution is much smaller than that of the antimode discussed above—the magnitude of the former, both absolutely and relative to the modes on either side, is much smaller than that of the latter.

Our findings here are consistent with, for example, those of Beaudry *et al.* (2005) who document increases in the 15–85 and smaller percentile ranges of the cross-country distribution of output per worker along with reductions in the 10–90 and larger percentile ranges between 1960 and 1998. They provide evidence that these changes began in the mid 1970s. Our statistical explanation of their findings is a tightening of the component distributions at either extreme of the cross-country distribution of output per worker at that time, which reduced the mass in the center of the distribution as well as in the tails.

To some extent, our results contrast with those of Sala-i-Martin (2006) who documents a small reduction in the dispersion of the world distribution of income between 1970 and 2000. However, his study and ours address two different issues as we follow the convergence literature and study the cross-country distribution of per capita GDP, using countries as our units of observation, whereas he studies the distribution of individual incomes, using people as his units of observation. Using a test based on kernel estimation methods, Henderson *et al.* (2008) find stronger evidence of multimodality using cross-country data weighted by population than with unweighted data similar to that used here.

More generally, in Section 3.5 we explain the often-discussed increase in the polarization of the cross-country distribution of output per worker since 1960 by an increase in the concentration of the poor and rich countries around their component means rather than by an increase in the gap between the means themselves.

3.3. *Mobility between Components*

As described above, we assign countries to components according to their maximal estimated conditional probability of belonging to each component. The $\hat{\zeta}_{ji}$ for each country, each component, and each year are available upon request. Given these assignments, we are able to observe the implied transitions between components that occur when assignments change. Given that the link between multimodality and the existence of convergence clubs is tenuous, this seems to be a more natural definition of a "transition" than the crossing from one side of an antimode to another (as used by, for example, Bianchi, 1997, and Henderson *et al.*, 2008). Moreover, it is not generally true that, if there is one, the antimode in a mixture distribution occurs at the point where the conditional probabilities of belonging to the two components are equal. That is, crossing from one side of the antimode to another need not imply a change in the component with the maximal conditional probability.

So defined, transitions are relatively rare events during our sample period and a small number of countries account for most of them so that immobility rather than mobility is the norm. Of the 816 possible transitions only 51, or about 6 percent, occur. Except for the flurry of transitions in the mid-1980s, the transition rate is roughly constant over the sample period. Sixty-four of the 102 countries in our sample remain assigned to the same component throughout the sample period.[10] Of those that do transit from their initial component, 28 shift just once so that the remaining 10 countries account for almost half of the observed transitions.

Of the countries that never leave their initial component, 18 are among the 26 initially rich countries, while the other 8 initially rich countries (Argentina, Chile,

[10]They are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Israel, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States (all initially in the rich group); Bolivia, Brazil, Columbia, the Dominican Republic, Ecuador, Egypt, El Salvador, Fiji, Gabon, Guatemala, Guyana, Honduras, Jamaica, Jordan, Malaysia, Morocco, Namibia, Nicaragua, Panama, Papua New Guinea, Paraguay, Philippines, Syria, and Turkey (all initially in the middle group); and Benin, Burkina Faso, Burundi, Chad, the Democratic Republic of the Congo, Ethiopia, the Gambia, Ghana, Guinea-Bissau, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Nepal, Niger, Nigeria, Rwanda, Tanzania, Togo, and Uganda (all initially in the poor group).

Costa Rica, Mexico, South Africa, Trinidad and Tobago, Uruguay, and Venezuela) move to the middle component where, with the exception of Argentina which returns to the rich component in 2000, they all remain until 2000.

Additionally, 24 of the countries that remain attached to the same component over the entire sample period are among the 40 countries initially classified as belonging to the middle component. Of the 16 countries that shift from the middle component during the sample period, 3 countries (Angola, Central African Republic, and Senegal) move to the poor component, 9 countries (Cyprus, Greece, Hong Kong, Japan, Korea, Mauritius, Portugal, Singapore, and Taiwan) move to the rich component, 2 countries (Cameroon and Guinea) return to it after spending the 1970s and 1980s in the poor component, and 2 countries (Iran and Peru) return to it after visiting the rich component in 1970.

The remaining 22 countries that stay attached to the same component over the entire sample period are among the 36 initially poor countries. Of the 14 countries that leave the poor component, 12 (Bangladesh, Botswana, China, Republic of the Congo, Cote d'Ivorie, India, Indonesia, Pakistan, Romania, Ski Lanka, Thailand, and Zimbabwe) move to the middle component and remain there until the end of the sample period, while 2 (Mauritania and Zambia) return to the poor component.

In sum, we conclude that the cross-component mobility during our sample period was low with transitions between components being relatively rare. While the transition rates that we find are low, they are somewhat higher than those documented in other studies such as Bianchi (1997), who finds that 3 of 238 (1.2 percent) possible transitions occur, Paap and van Dijk (1998), who find that 21 of 720 (2.9 percent) possible transitions occur, and Henderson *et al.* (2008), who find that 12 of 291 (4.1 percent) and 19 of 414 (4.6 percent) possible transitions occur in the two per capita output datasets that they employ.[11] Our higher estimated transition rate is explained, in part at least, by our greater number of putative convergence clubs. Each of the studies cited above identifies two putative clubs, whereas we find three so that we have twice as many between-club boundaries and hence twice as many points where a transition can occur. We would thus expect to observe a higher frequency of transitions given any degree of mobility within the distribution. Nonetheless, we conclude that the mobility between the components of the cross country distribution of per capita income is low.[12]

In addition to allocating the countries among the components, the estimated conditional probabilities can be used to measure the strength of the affinity between countries and components. Higher probabilities indicate tighter bonds, so we measure the overall tightness of the bonds between countries and components by counting the number of countries with a $\hat{\zeta}_{ji}$ greater than a threshold, τ, for any j. Figure 4 plots the number of countries with a $\hat{\zeta}_{ji} > \tau$ for any j for various values

[11]Henderson *et al.* (2008) follow Bianchi (1997) and define transitions as movements across the antimode between the two modes that they identify, while Paap and van Dijk (1998) do as we do and count changes in component assignments based on maximal estimated conditional probabilities.

[12]In treating a change in the component with the maximal conditional probability as a transition between components we are ignoring the possibilities of a more accurate assignment to the correct component or a less accurate assignment to an incorrect component. To this extent our approach tends to overstate the amount of mobility in the sample.

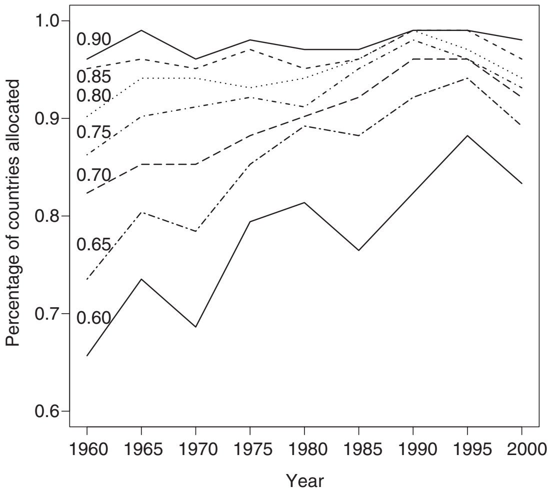

Figure 4. Percentage of Countries Allocated for Different Values of the Threshold

of τ from 0.6 to 0.9. The first feature evident in Figure 4 is the general strength of the bonding between countries and components. In 1960, for example, about two-thirds of the countries have a $\hat{\zeta}_{ji} > 0.9$ for some j and over 90 percent have a $\hat{\zeta}_{ji} > 0.7$. The second, and arguably more important, feature displayed in Figure 4 is the evident increase in the strength of the affinity between countries and components over the sample period as shown by the general tendency for the lines in Figure 4 to rise over time. For example, in 1960, 74 percent of the countries had a maximal conditional probability greater than 0.85, whereas in 2000, 89 percent were bound this tightly to a component. The tendency for the number of countries with a $\hat{\zeta}_{ji} > \tau$ to rise between 1960 and 1998 holds for all values of τ although it is necessarily less pronounced for lower values. In sum, we conclude that most countries are bound very tightly to a component and that the tightness of the bonds has increased over time.

Our finding of low cross-component mobility leads us to interpret the multiple mixture components identified in Section 3.1 as representing multiple basins of attraction in the stochastic process describing the evolution of output per worker.[13] That is, we regard that process as characterized by convergence clubs so that a country's initial level of output per worker plays an important role in determining its long-run level. Moreover, the role of initial conditions seems to be strengthening as the affinity of countries for clubs became stronger during the period that we have studied. It is important to note at this point that our results are subject to a version of the identification caveat discussed in Durlauf and Johnson (1995). The behavior that we have documented is compatible with a model in which there are multiple steady states, or convergence clubs, as we have emphasized, as well as

[13]Quah (1996, 1997) measures intradistributional immobility by the concentration of mass along the main diagonal of the stochastic kernels that he estimates. That immobility is what we have called "low cross-component mobility" and so countries having high estimated ζ_{ji} values for the same j over time are those that would be found at places in the stochastic kernels where the mass is concentrated along the main diagonal.

with a model in which countries transit through different stages of development before reaching a common (stochastic) steady state. In common with all of the empirical growth literature, differentiation between these two alternatives is hampered by the time span of our dataset.

3.4. *Behavior within Components*

Having discussed the mobility within the distribution we return now to the issue of the role of compositional changes in the reduction of the variance of the rich component over the sample period. Recall that this reduction occurs mainly in two steps, viz. the fall between 1970 and 1975 and that between 1985 and 1990. While the latter is due in some part to the movement of Argentina, Mexico, South Africa, Trinidad and Tobago, and Venezuela out of the group, the role of such compositional changes in the former is small, as it is in the total reduction in the variance of the rich component over the sample period. To show this we consider the variation in output per worker across the 18 countries that remain assigned to the rich component throughout the sample period. Figure 7 plots the standard deviation of output per worker across this "always rich" group as well as that for rich component. Over the sample period the former fell by almost two-thirds, with about 75 percent of the decline occurring before 1975. By contrast, the estimated standard deviation of the rich component falls by about 50 percent over the sample period. This implies that the observed tendency for the rich component to become increasingly separated from the other two components is not due to compositional effects but rather to forces within the group causing the rich countries to become increasingly concentrated around the group mean. That is, the rich component did not become more concentrated around its mean simply because some countries relatively far from the mean left the group. Instead, the rich countries tended to move closer to each other and in doing so increased their separation from the other countries in the world.[14]

Figures 5 and 6 show analogous information for the poor and middle components, respectively. As with the rich component, in both of these cases the behavior of the standard deviations of the group of countries always assigned to each component mirrors that of the corresponding estimated component standard deviation. Figure 5 shows that, as with the estimated standard deviation of the poor component, with the exception of the late 1970s, the standard deviation of output per worker in the 22 "always-poor" countries fell steadily from 1960 to 2000. As with the rich component, the poor component did not become more concentrated around its mean simply because some countries relatively far from the mean left the group. Rather, the poor countries tended to move closer to each other and in doing so also increased their separation from the other countries in the world. Figure 6 shows that both the estimated standard deviation of the middle component and the standard deviation of the income per capita across the 24 "always middle" countries exhibit a slight upward trend over the 1960 to 2000 period. Note that despite our attempt to control for the effect of countries leaving

[14]These results suggest the possibility of σ-convergence within each of the poor and rich groups. See Durlauf *et al.* (2005) for a discussion of various modes of convergence, and Young *et al.* (2008) as well as Egger and Pfaffermayr (2009) for recent studies that employ this concept.

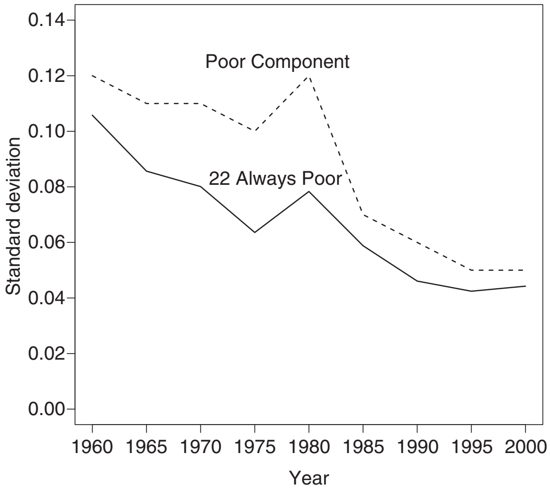

Figure 5. Standard Deviation of the Poor Group

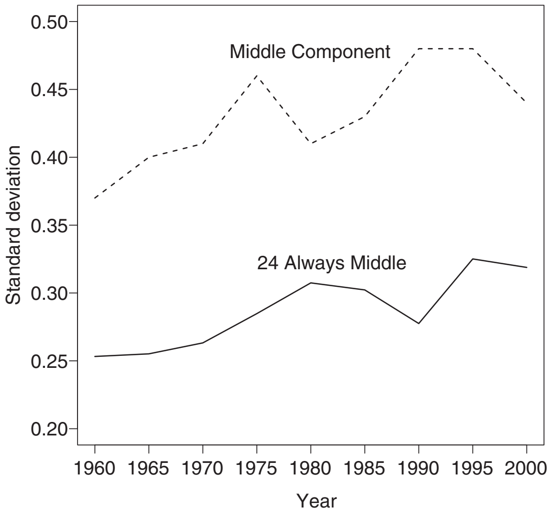

Figure 6. Standard Deviation of the Middle Group

and entering the middle group, the caveat discussed at the end of the previous section implies that we cannot rule out the possibility that the middle group will eventually disappear as its members jump to either the poor or the rich group. The increase in the variance within this group could be the precursor to these events.

3.5. *Evolution of Inequality and Polarization*

The evolution of the cross-country distribution of per capita income that we document above has implications for the degrees of inequality and polarization of the distribution. One way to formalize these implications is to compute the

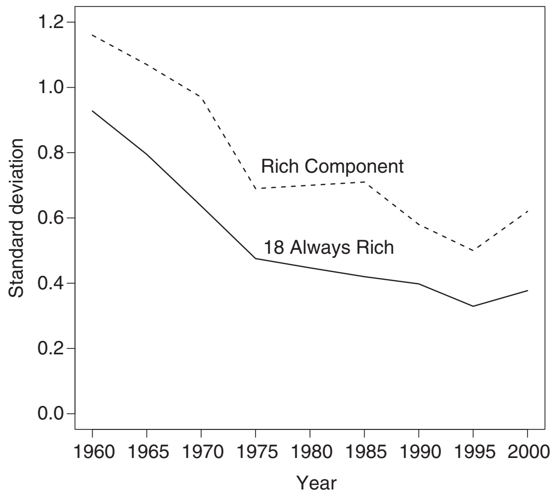

Figure 7. Standard Deviation of the Rich Group

TABLE 3
INEQUALITY AND POLARIZATION MEASURES

	$\alpha = 0$			$\alpha = 0.25$			$\alpha = 0.5$			$\alpha = 0.75$			$\alpha = 1$		
	P_α	P_α^*	P_α^σ	P_α	P_α^*	P_α^σ	P_α	P_α^*	P_α^σ	P_α	P_α^*	P_α^σ	P_α	P_α^*	P_α^σ
1960	1.09	1.09	1.09	0.79	0.79	0.79	0.61	0.61	0.61	0.50	0.50	0.50	0.43	0.43	0.43
1965	1.12	1.09	1.12	0.82	0.80	0.82	0.64	0.63	0.64	0.54	0.53	0.54	0.47	0.47	0.48
1970	1.15	1.16	1.13	0.84	0.85	0.83	0.66	0.67	0.65	0.56	0.56	0.54	0.49	0.50	0.47
1975	1.17	1.06	1.13	0.87	0.80	0.85	0.69	0.64	0.68	0.58	0.54	0.58	0.51	0.49	0.52
1980	1.14	1.11	1.12	0.85	0.83	0.83	0.67	0.66	0.65	0.55	0.55	0.54	0.48	0.48	0.46
1985	1.12	1.09	1.17	0.83	0.82	0.87	0.66	0.65	0.73	0.55	0.55	0.65	0.49	0.48	0.62
1990	1.15	1.04	1.13	0.87	0.79	0.89	0.71	0.65	0.77	0.62	0.57	0.73	0.58	0.54	0.74
1995	1.11	1.05	1.12	0.86	0.82	0.90	0.72	0.69	0.81	0.65	0.63	0.80	0.63	0.61	0.84
2000	1.11	1.10	1.13	0.84	0.85	0.90	0.70	0.71	0.80	0.63	0.64	0.78	0.61	0.61	0.82

Notes: P_α denotes the Duclos, Esteban, and Ray index of polarization for a range of values of the parameter α. P_α for $\alpha = 0$ is equivalent to twice the Gini index of inequality. P_α^* measures the polarization with the component means held fixed at the estimated 1960 values. P_α^σ measures the polarization with the component means and mixing proportions held fixed at the estimated 1960 values.

polarization measure proposed by Duclos, Esteban and Ray (2004). For a population with income distribution described by the density $f(x)$, this measure is:

$$P_\alpha(f) = \iint f(x)^{1+\alpha} f(y)|y-x|\,dydx \tag{5}$$

where $\alpha \in [0.25, 1]$ is a parameter that indexes the identification effect in the identification–alienation framework used by the authors. As they point out, $P_0(f)$ is twice the Gini coefficient although this value of α lies below the lower bound of 0.25 implied by their axioms. Table 3 shows estimates of $P_\alpha(f)$ for $\alpha = 0, 0.25, 0.5, 0.75, 1$, computed with $f(x)$ replaced by the estimated three-component mixture model of the cross-country distribution of per capita income for each of our sample years between 1960 and 2000. These measures indicate that, over this period, the inequality in the distribution, as measured by (twice) the Gini coefficient, $P_0(f)$, has fluctuated somehow, rising in the late 1960s and then falling in the early 1990s to finish the period virtually unchanged. In other words, at least as measured by the Gini coefficient, measured inequality in 2000 was about the same as it was in 1960. In contrast, for each value of $\alpha \geq 0.25$ that we use, measured polarization rises over this period. Both the absolute and proportional rises are increasing in α and, with the exception of the late 1970s and the late 1990s, these rises are monotonic. As we show below, this rise is driven by the tightening of the rich and poor component distributions around their respective means. The decreased dispersion within these distributions tends to increase the within-club "identification" and so measured polarization because of the weight given to this effect.[15]

To study the statistical causes of the rise in polarization we compute $P_\alpha^*(f)$, which is a version of the computed $P_\alpha(f)$ with the component means held fixed at their estimated 1960 values. A comparison of $P_\alpha^*(f)$ and $P_\alpha(f)$ thus enables us to gauge the role of the changing component means in the rise in polarization. As noted in Section 3.2, the gap between the rich and poor component means widens over that sample period—a phenomenon that would tend to increase inequality and polarization. We also compute $P_\alpha^\sigma(f)$, which is a version of the computed $P_\alpha(f)$ with both the component means and the mixing proportions held fixed at their estimated 1960 values. Changes in $P_\alpha^\sigma(f)$ thus reflect only the effects of the changes in the component standard deviations and a comparison of $P_\alpha^\sigma(f)$ and $P_\alpha^*(f)$ isolates the effects of changes in the mixing weights. As noted in Section 3.2, the standard deviations of the rich and poor components have fallen substantially over the sample period while that of the middle component has risen somewhat—phenomena that together would have ambiguous effects on polarization and inequality. Both $P_\alpha^*(f)$ and $P_\alpha^\sigma(f)$ are shown alongside $P_\alpha(f)$ in Table 3 for each value of α and for each year. Additionally, Figure 8 plots $P_{0.75}(f)$, $P_{0.75}^*(f)$, and $P_{0.75}^\sigma(f)$ against time.

In the case of the inequality measures, $P_0^*(f)$ fluctuates less than $P_0(f)$ and not always in the same direction—in both 1975 and 1990, $P_0(f)$ rises while $P_0^*(f)$ falls sharply. These two years saw large rises in $\hat{\mu}_{\text{rich}}$ which increased the gaps between it and $\hat{\mu}_{\text{poor}}$ and $\hat{\mu}_{\text{middle}}$ by about 12 percent in each case—the largest changes in these gaps that we observe. This suggests that the changes of the component means can be an important source of the variation in inequality. However, despite the rises in both $\hat{\mu}_{\text{rich}} - \hat{\mu}_{\text{poor}}$ and $\hat{\mu}_{\text{rich}} - \hat{\mu}_{\text{middle}}$ over the entire sample period noted in Section 3.2, measured inequality is virtually the same at the end of the period as at the beginning.

[15]These results accord with those of Anderson (2004), who uses stochastic dominance concepts to study inequality and polarization in the cross-country distribution of per capita GNP over the 1970 to 1995 period. We do not apply his methods as they require constancy of the relative club sizes which is not the case here. A measure of overlapping of the distributions which is directly related to the assessment of polarization between groups could be estimated. A non-parametric estimator based on kernel density estimates has been proposed in Anderson *et al.* (forthcoming), and more recently further analyzed in Anderson *et al.* (2009).

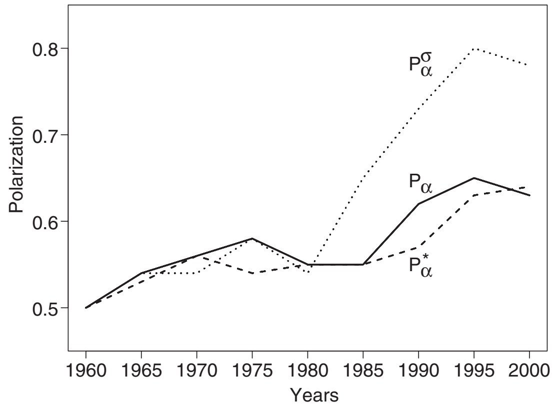

Figure 8. Trend of Polarization (DER, $\alpha = 0.75$)

Moreover, in 2000, $P_0^\sigma(f)$ is slightly greater than both $P_0^\sigma(f)$ in 1960 and $P_0(f)$ in 2000, showing that the net effect of the changes in the component standard deviations is also to (slightly) increase inequality. The reason that the rise in inequality is much less than the increased gaps between $\hat{\mu}_{\text{rich}}$ and $\hat{\mu}_{\text{poor}}$ and $\hat{\mu}_{\text{middle}}$ alone would imply is the offset provided by the changes in the estimated mixing proportions, the $\hat{\pi}_j$. While $\hat{\pi}_{\text{rich}}$ is relatively constant at about 0.27 throughout the sample period, the behavior of $\hat{\pi}_{\text{middle}}$ resembles a step function with a jump from about 0.4 to about 0.5 between 1980 and 1985, while $\hat{\pi}_{\text{poor}}$ exhibits a corresponding fall from around 0.33 to 0.23. This large shift in mass toward the middle component dramatically reduces measured inequality.

For each value of $\alpha \geq 0$, the behavior of $P_\alpha^*(f)$ in most years is very similar to that of $P_\alpha(f)$, indicating that the net effect of the changes in the component means on measured polarization is small. This similarity tends to increase with α in that the differences between $P_\alpha^*(f)$ and $P_\alpha(f)$ decline as α rises.

As higher values of α increase the weight given to within-club "identification" this implies that it is increases in that aspect of the polarization measure that is at work here—a claim that is consistent with the differences in the behaviors of the inequality and polarization measures.

The computed $P_\alpha^\sigma(f)$ measures rise steadily through the sample period. Until 1980, they track the corresponding $P_\alpha(f)$ and $P_\alpha^*(f)$ measures very closely, implying that changes in the $\hat{\sigma}_j$ are primarily responsible for the rise in polarization from 1960 to 1980. As Figure 8 illustrates for $\alpha = 0.75$, after 1980 the paths of $P_\alpha^\sigma(f)$ and the other two polarization measures diverge, with $P_\alpha^\sigma(f)$ rising more quickly than the others. The divergence between $P_\alpha^\sigma(f)$ and $P_\alpha^*(f)$ implies that, while changes in the $\hat{\sigma}_j$ remain an important factor in the rise in polarization after 1980, some counteracting effect is provided here by the changes in the estimated mixing proportions detailed above. The large shift in mass toward the middle component tends to reduce polarization and opens the gap between $P_\alpha^\sigma(f)$ and $P_\alpha^*(f)$ evident from 1985 onwards.

In sum, while inequality in the cross-country definition of per capita income, as measured by the Gini coefficient, is about the same in 2000 as it was in 1960, albeit with some fluctuations, the polarization in the distribution, at least as measured by $P_\alpha(f)$, rises steadily from 1960 to 2000. The primary proximate cause of this rise is the narrowing of the rich and poor component distributions. As the countries in the convergence clubs represented by those components become more concentrated around their respective club means, they become more like each other and less like the countries in other convergence clubs. This increases cross-country polarization in the overall distribution.

4. Conclusions

We have argued that, despite the attention that it has received in the literature, multimodality of the cross-country distribution of per capita output is neither necessary nor sufficient for the presence of multiple basins of attraction, or convergence clubs, in the dynamic process describing the evolution of that distribution over time. Kernel estimation methods and the associated "bump hunting" approaches to the detection of multiple modes are thus likely to be less informative, when investigating the convergence hypothesis, than approaches which model the distribution as a mixture of component densities. Each of these densities represents a putative convergence club and the mixture approach provides integrated tests for number of components. As tests of the convergence hypothesis, these tests have greater power than multimodality tests because the multiple components may not reveal themselves as multiple modes. Moreover, the estimated *ex post* probabilities that a country belongs to each of the components provides a natural metric for assigning countries to components and, more generally, for measuring the strength of the affinity between countries and components. Comparison of such assignments over time provides a natural framework for the assessment of mobility between components, which is important as low mobility is an essential part of a convergence club view. Even if multiple components are detected, high mobility between them would be contrary to the claim that they represented a multiple basis of attraction.

We implement the mixture approach using cross-country per capita income data for the period 1960 to 2000. In contrast to the commonly held view that the cross-country distribution of per capita income exhibits two modes, both of the statistical tests that we use indicate the presence of three-component densities in each of the nine years that we examine over this period. For each year we thus estimate a three-component mixture model and label the components as "poor," "middle," and "rich." We find that, while the gap between the mean relative per capita incomes of the rich and poor group has widened somewhat, the evident "hollowing out" of the middle of the distribution is largely attributable to the increased concentration of the rich and poor countries around their respective component means. This explanation is robust to the compositional changes brought by the few transitions out of these two groups that do occur. We track those transitions by using the estimated *ex post* probabilities of component membership to assign each country to a component in each year. While transitions do occur, they are rare, with only about 7 percent of the possible transitions actually occurring. Of the 102 countries in our sample, 64 remain assigned to the same component throughout the sample period and 28 transit just once, so that the remaining 10 countries account for over 40 percent of the observed transitions. This finding of low cross-component mobility leads us to interpret the multiple mixture components that we detect as representing convergence clubs.[16] There is a pronounced tendency for the maximal *ex post* probability for each country to increase, indicating a strengthening of the affinity of countries for the club in which they lie. Finally, we use our estimated mixture densities to compute measures of polarization and find that they have increased over the sample period—a phenomenon that we attribute primarily to the decreased variances of the poor and rich components.

References

Anderson, Gordon, "Making Inferences About the Polarization, Welfare, and Poverty of Nations: A Study of 101 Countries 1970–1995," *Journal of Applied Econometrics*, 19, 537–50, 2004.

Anderson, Gordon, Oliver Linton, and Yoon-Jae Wang, "Non Parametric Estimation of a Polarization Measure," Working Paper 363, Department of Economics, University of Toronto, July 2009.

Anderson, Gordon, Ying Ge, and Teng Wah Leo, "Distributional Overlap: Simple, Multivariate, Parametric and Nonparametric Tests for Alienation, Convergence and General Distributional Difference Issues," *Econometric Reviews*, forthcoming.

Azariadis, Costas, "The Economics of Poverty Traps: Part One: Complete Markets," *Journal of Economic Growth*, 1, 449–96, 1996.

Beaudry, Paul, Fabrice Collard, and David A. Green, "Changes in the World Distribution of Output-per-capita 1960–98: How a Standard Decomposition Tells an Unorthodox Story," *The Review of Economics and Statistics*, 87, 741–53, 2005.

Bianchi, Marco, "Testing for Convergence: Evidence from Nonparametric Multimodality Tests," *Journal of Applied Econometrics*, 12, 393–409, 1997.

Bloom, David E., David Canning, and Jaypee Sevilla, "Geography and Poverty Traps," *Journal of Economic Growth*, 8, 355–78, 2003.

Chen, Hanfeng, Jiahua Chen, and John D. Kalbfleisch, "Testing for a Finite Mixture Model with Two Components," *Journal of the Royal Statistical Society B*, 66, 95–115, 2004.

Chen, Jiahua and John D. Kalbfleisch, "Modified Likelihood Ratio Test in Finite Mixture Models with a Structural Parameter," *Journal of Statistical Planning and Inference*, 129, 93–107, 2005.

Davis, Lewis S., Ann L. Owen, and Julio Videras, *Do All Countries Follow the Same Growth Process?* (http://ssrn.com/abstract=1064521), 2007.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin, "Maximum Likelihood from Incomplete Data via EM Algorithm," *Journal of the Royal Statistical Society B*, 69, 1–38, 1977.

Duclos, Jean-Yves, Joan Esteban, and Debraj Ray, "Polarization: Concepts, Measurement, Estimation," *Econometrica*, 72, 1737–72, 2004.

Durlauf, Steven N. and Paul A. Johnson, "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics*, 10, 365–84, 1995.

Durlauf, Steven N., Paul A. Johnson, and Jonathan Temple, "Growth Econometrics," in Philippe Aghion and Steven N. Durlauf (eds), *Handbook of Economic Growth*, North-Holland: Elsevier, Amsterdam, 1063–114, 2005.

Egger, Peter and Michael Pfaffermayr, "On Testing Conditional Sigma-Convergence," *Oxford Bulletin of Economics and Statistics*, 71, 453–73, 2009.

[16]Taken as test of the convergence hypothesis, our paper has studied "unconditional" convergence in that we do not introduce variables to control for the cross-country heterogeneity that may influence long-run behavior. See Durlauf *et al.* (2005) for more on the difference between "conditional" and "unconditional" convergence. A sequel will allow for "conditional" convergence by parameterizing the mixture model parameters as functions of relevant control variables. In a related paper, Davis *et al.* (2007) have estimated a finite mixture model of the cross-country distribution of the growth rate of real GDP (conditional on the Mankiw *et al.* (1992) control variables) that allows the mixing proportions to depend on several covariates. Consistent with the results of Durlauf and Johnson (1995) and others, they find evidence of multiple growth regimes.

Fan, Yanqin, "Bootstrapping a Consistent Non Parametric Goodness-of-Fit Test," *Econometrics Reviews*, 14, 367–82, 1995.

Galor, Oded, "Convergence? Inferences from Theoretical Models," *The Economic Journal*, 106, 1056–69, 1996.

Ghosh, Jayanta K. and Pranab K. Sen, "On the Asymptotic Performance of the Loglikelihood Ratio Statistic for the Mixture Model and Related Results," in Lucien Le Cam and Richard A. Olshen (eds), *Proceedings of the Berkely Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. 2, Wadsworth, Monterey, 789–806, 1985.

Henderson, Daniel J., Christopher Parmeter, and R. Robert Russell, "Modes, Weighted Modes, and Calibrated Modes: Evidence of Clustering Using Modality Tests," *Journal of Applied Econometrics*, 23, 607–38, 2008.

Heston, Alan, Robert Summers, and Bettina Aten, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP) (http://pwt.econ.upenn.edu/), 2002.

Holzmann, Hajo, Sebastian Vollmer, and Julian Weisbrod, *Twin Peaks or Three Components? Analyzing the World's Cross-Country Distribution of Income*, IAI Discussion Paper No.162, University of Gottingen (http://www2.vwl.wiso.uni-goettingen.de/ibero/papers/DB162.pdf), 2007.

Jones, Charles I., "On the Evolution of the World Income Distribution," *Journal of Economic Perspectives*, 11, 19–36, 1997.

Jones, M. Chris, J. Steve Marron, and Simon J. Sheather, "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, 91, 401–7, 1996.

Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107, 407–37, 1992.

Marron, J. Steve and Matt P. Wand, "Exact Mean Integrated Squared Error," *The Annals of Statistics*, 2, 712–36, 1992.

McLachlan, Geoffrey, "On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture", *Journal of the Royal Statistical Society C*, 36, 318–24, 1987.

McLachlan, Geoffrey and David Peel, *Finite Mixture Models*, Wiley, New York, 2000.

Paap, Richard and Herman K. van Dijk, "Distribution and Mobility of Wealth of Nations," *European Economic Review*, 42, 1269–93, 1998.

Pittau, Maria Grazia, "Fitting Regional Income Distributions in the European Union," *Oxford Bulletin of Economics and Statistics*, 67, 135–61, 2005.

Pittau, Maria Grazia and Roberto Zelli, "Empirical Evidence of Income Dynamics Across EU Regions", *Journal of Applied Econometrics*, 21, 605–28, 2006.

Quah, Danny, "Convergence Empirics Across Economies with (Some) Capital Mobility," *Journal of Economic Growth*, 1, 95–124, 1996.

———, "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs," *Journal of Economic Growth*, 2, 27–59, 1997.

Sala-i-Martin, Xavier, "The World Distribution of Income: Falling Poverty and . . . Convergence, Period," *Quarterly Journal of Economics*, 121, 351–97, 2006.

Sheather, Simon J. and M. Chris Jones, "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation," *Journal of Royal Statistical Society B*, 53, 683–90, 1991.

Silverman, Bernard W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

Tsionas, Efthymios G., "Regional Growth and Convergence: Evidence from the United States," *Regional Studies*, 34, 231–8, 2000.

Young, Andrew T., Matthew John Higgins, and Daniel Levy, "Sigma Convergence versus Beta Convergence: Evidence from U.S. County-Level Data," *Journal of Money, Credit and Banking*, 40, 1083–93, 2008.