Vollmer, Sebastian; Holzmann, Hajo; Weisbrod, Julian

Working Paper

Peaks vs. components

Diskussionsbeitrag, No. 452

Provided in Cooperation with:
School of Economics and Management, University of Hannover

Suggested Citation: Vollmer, Sebastian; Holzmann, Hajo; Weisbrod, Julian (2010) :
Peaks vs. components, Diskussionsbeitrag, No. 452, Leibniz Universität Hannover,
Wirtschaftswissenschaftliche Fakultät, Hannover

This Version is available at:
https://hdl.handle.net/10419/66012

WWW.ECONSTOR.EU

Mitglied der

Leibniz-Gemeinschaft

Peaks vs. Components[*]

Sebastian Vollmer, Harvard University & University of Hannover
Hajo Holzmann, University of Marburg
Julian Weisbrod, University of Goettingen

Abstract

We analyze the cross-national distribution of GDP per capita and its evolution from 1970 to 2003. We argue that peaks are not a suitable measure for distinct growth regimes, because the number of peaks is not invariant under strictly monotonic transformations of the data (e.g. original vs. log scale). Instead, we model the distribution as a finite mixture, and determine its number of components (and hence of distinct growth regimes) from the data by rigorous statistical testing. We find that the distribution appears to have only two components in 1970-1975, but consists of three components from 1976 onwards. The level of GDP per capita stagnated in the poorest component, and the richest component grew much faster than the medium component. These findings empirically confirm the predictions of the unified growth theory.

JEL classification: C12, O11, O47, F01
Keywords: twin peaks, economic growth, convergence.

---

[*]This paper is a substantially revised version of "Twin Peaks or Three Components? - Analyzing the Cross-Country Distribution of Income" *Ibero-America Institute for Economic Research Discussion Paper* 162.

1

1 Introduction

The notion of twin peaks in the cross-country income distribution was introduced by Quah (1993, 1996, 1997). He interpreted the emergence of twin peaks as polarization of the cross-country income distribution into a rich and a poor convergence club. Bianchi (1997) confirmed Quah's observation of twin peaks via rigorous statistical testing. The contributions of Quah are part of a larger literature on convergence (e.g. Barro, 1991; Barro and Sala-i-Martin, 1992; Mankiw, Romer and Weil, 1992; Sala-i-Martin, 1996; Galor, 1996; Jones, 1997; Graham and Temple, 2006). It is controversial whether the twin peaks represent locally stable equilibria/convergence clubs (Quah, 1996) or whether they are only a temporary phenomenon due to a high frequency of growth miracles (Jones, 1997).

In this paper we will challenge the twin peaks result. It turns out that the number of peaks of a distribution is not preserved under strictly monotonic transformations of the data. A simple log transformation will change the number of components in the cross-country income distribution from two to three (as we will see later). Although this result is not obvious at all, the intuition is rather straightforward: Compression/decompression of the data causes the different modes to merge/separate. This property makes peaks practically useless for economic interpretations. It certainly doesn't make sense to assign countries to the middle-income group on the log scale and to the low-income group on the original scale. It also doesn't make sense to interpret peaks as convergence clubs, since a simple log transformation can make them go away or emerge.

We propose a method to identify different regimes within a distribution, which is

not affected by the scale of the data. With this method we find that the cross-country distribution of income consists of three statistically distinguishable regimes from the mid 1970s onwards, in 1970 we can only identify two regimes. Pittau et al. (2010) apply a similar methodology to the cross-country income distribution and analyze and interpret their empirical results in great detail. However, they firstly do not make the important point that modes are scale-dependent, and thus components are clearly superior for economic interpretations. Secondly, we apply the new methodology of modified likelihood ratio tests for homogeneity in finite mixtures to the problem, rather than simply bootstrapping the likelihood ratio statistic without theoretical support. Finally, our interpretation and framing of the results differ from Pittau et al. (2010).

The unified growth theory (c.f. Galor, 2009 for an overview) provides another explanation for multiple regimes which also uncovers the forces that have lead to the emergence of these regimes. The theory suggests that growth segments economies into three fundamental regimes: a malthusian regime with slow growing economies, fast growing economies in a sustained growth regime, and a third group in the transition from one regime to the other. One important difference to models with multiple equilibria is that this segmentation does not represent the long-run steady state of these economies. Variations in the levels of income only reflect country-specific characteristics and not the actual stage of development. Thus, there are no critical levels that permit economies to switch from one regime to the other, but rather critical rates of progress. We will see that our empirical findings support the predictions of the unified growth theory.

2 Peaks

We use GDP per capita data (PPP, chain series with base year 2000) from the Penn World Tables 6.2 (Summer, Heston & Aten, 2006). In order to compare our observations over time, we restrict ourselves to those countries, of which complete income data for the whole time period are available.[1] This leaves 124 countries for the period from 1970 to 2003 in the sample.
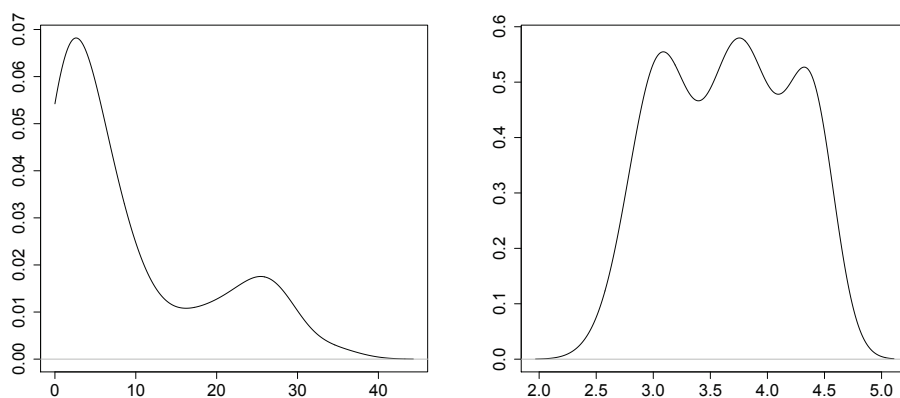
Figure 1: Kernel density estimate for GDP per capita (left) and log GDP per capita (right) in 2003. For conveniently interpreting the figure, we here use the logarithm to the base 10

Figure 1 shows simple kernel density estimates for our data in 2003 on the original scale ($1000) and on a log scale. From visual inspection alone we can see that the density of the original data has two peaks while it has three peaks on the log scale. We

---

[1] Further we exclude countries with less than half a million inhabitants. The GDP per capita of these countries is very sensitive towards economic shocks (e.g. oil) and doesn't necessary reflect their fundamental economic development. This restriction doesn't affect our main point on peaks versus regimes, but it should help provide a clearer picture of the world's cross-country distribution of income.

have chosen the year 2003 because the difference between the original scale and the log scale is very pronounced. One reason why Quah and others might have overseen this problem could be that the third peak in the log density wasn't as pronounced in the 1970s and 1980s.

The different numbers of peaks in the plots could be a simple artifact of the nonparametric curev etimates, e.g. from inaccurate choice of the tuning parameter. It is therefore necessary to validate the statistical significance of the peaks via rigorous statistical testing. To this end we utilize Silverman's test. Formally, a *peak* of a density f (and similarly of the kernel estimator $\hat{f}$) is a local maximum of f (or $\hat{f}$). Silverman (1981) showed that the number of modes of $\hat{f}$ is a right-continuous, monotonically decreasing function of the bandwidth h if the normal kernel $K(x) = (2\pi)^{-1}\exp(-x^2/2)$ is employed. This allowed him to define the k-critical bandwidth $h_c(k)$ as the minimal bandwidth h for which $\hat{f}$ still just has k modes and not yet $k+1$ modes. Based on the notion of the k-critical bandwidth, Silverman (1981) proposed a bootstrap test for the hypothesis

$$\tilde{H}_k : f \text{ has at most } k \text{ modes} \qquad \text{against} \qquad \tilde{K}_k : f \text{ has more than } k \text{ modes.}$$

We apply this test to the GDP per capita and log-GDP per capita data for the year 2003. We can clearly reject the hypothesis $\tilde{H}_2$ of two peaks in favor of the alternative of more than two peaks for the log data (p-value< 0.001). For the data on the original scale we can reject the hypothesis $\tilde{H}_1$ with a p-value of < 0.001, but we cannot reject the hypothesis $\tilde{H}_2$ with a p-value of about 0.45. We thus conclude that the number of peaks which is visible in the plots (two on the original scale and

5

three on the log scale) is indeed statistically significant. With this simple exercise we have shown that the number of peaks of a density depends on the scaling of the data. Thus, it should not be used for any interpretations on convergence clubs.

3 Components

Let f denote the density of the cross-country distribution of GDP per capita for a given year. We model f as a finite mixture

$$f(x) = \alpha_1 g(x; \phi_1) + \ldots + \alpha_m g(x; \phi_m), \qquad x > 0, \tag{1}$$

where $g(x; \phi)$ is a parametric family of densities and the weights $\alpha_i \geq 0$ sum up to one. There is no general simple connection between the number of modes of f and the number of components m. Typically, for unimodal g, the number of modes of f will be at most m, but often will be less than m. The number of components is preserved if the data are transformed via a strictly monotonic transformation (if candidate densities are correspondingly transformed).

Paap and Dijk (1998) also used a mixture to model the cross-country distribution of GDP per capita. However, their model resembles the fit of a histogram, thus, the "stylized fact" of a distinction between poor and rich countries is already built into their model. We believe that the data itself should determine the number of components via statistical inference. To this end a finite mixture with normal components of the log-income distribution is the appropriate tool.[2]

---

[2]   We restrict the model class of finite normal mixtures to have equal variances. There are

Testing in parametric models is often accomplished by using the likelihood ratio test (LRT). However, the standard theory of the LRT does not apply for the number of components in finite mixture models (Dacunha-Castelle and Gassiat, 1999). Recently, Chen et al. (2001, 2004) and Chen and Kalbfleisch (2005) suggested modified LRTs, which retain a comparatively simple limit theory as well as the good power properties of the LRT. We shall apply these tests to our problem concerning the number of components in the distribution. At this point, we want to mention that the LRT and also the modified LRT are invariant under strictly monotonic transformation of the data. Thus, we could test on the original scale and on the log scale, the results are (in contrast to Silverman's test) perfectly consistent. For convenience, we shall use the log-data. Apart from testing the number of components, we also compare the mixture models via two popular model selection criteria, namely the Akaike information criterion (AIC, c.f. Akaike, 1978) and the Bayesian information criterion (BIC, Schwarz, 1978), given by $-2l + 2k$ and $-2l + k \log n$, respectively, where l is the log-likelihood, k the number of parameters and n the number of observations.

We first consider testing one against two components in a mixture. Suppose that $\phi(y; \mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ, and

two main reasons for this restriction: First, the likelihood function is unbounded in mixtures of normal distributions with distinct variances. Second, if distinct variances are allowed, the posterior analysis is no longer consistent. For example, the parameter estimate of the standard deviation of the richest component is about twenty times smaller than the standard deviation of the other components in a model with unequal variances for 2003. The U.S. would not be assigned to the richest component in this case. Our model can describe the data as good as the model with distinct variances, but it is not affected by these shortcomings of the more general model. Moreover, we would like to note that the use of the model with distinct variances would not change anything about the conclusion on how many components are required.

consider the two-component mixture

$$f(y; \alpha, \mu_1, \mu_2, \sigma) = \alpha\phi(y; \mu_1, \sigma) + (1 - \alpha)\phi(y; \mu_2, \sigma) \qquad (2)$$

with equal standard deviation σ. The testing problem is

$$H_1 : \quad f \text{ is normally distributed} \qquad \text{against} \qquad K_1 : \quad f \text{ is of the form (2)}.$$

Chen, Chen and Kalbfleisch (2001) show that for known σ, the likelihood-ratio statistic asymptotically follows the distribution $1/2\chi_0^2 + 1/2\chi_1^2$, where χ_0^2 is the point mass at zero.

Chen, Chen and Kalbfleisch (2004) also consider the problem of testing for two against more components of a mixture distribution. More precisely, the problem is to test

$$H_2 : f \text{ is of the form (2)} \qquad \text{against} \qquad K_2 : \quad f \text{ has more than two components.}$$

They show that given a known σ, the modified likelihood-ratio statistic is asymptotically distributed as $q\chi_0^2 + \frac{1}{2}\chi_1^2 + (1 - q)\chi_2^2$, where the proportion q depends on the mixing distributions.

Table 1 displays the results of the modified likelihood ratio test for one vs. two components and two vs. three components as well as the AIC and BIC model selection criteria for the respective fitted models. First of all, we note that two components are always preferable to one. In 1970 we cannot reject the hypothesis of

8

	One Component		1 vs. 2	Two Components		2 vs. 3	Three Components		Four Components	
Year	AIC	BIC	p-value	AIC	BIC	p-value	AIC	BIC	AIC	BIC
1970	152.5	158.1	< 0.01	142.6	153.9	0.87	145.9	162.8	145.3	167.8
1974	159.3	164.9	< 0.01	149.6	160.9	0.24	149.5	166.4	149.5	172.0
1975	157.5	163.1	< 0.01	147.5	158.8	0.17	146.5	163.4	146.4	169.0
1976	159.7	165.4	< 0.01	149.7	161.0	0.05	145.8	162.7	145.4	168.0
1980	167.2	172.8	< 0.01	155.7	167.00	0.04	151.2	168.1	155.2	177.8
1985	173.1	178.7	< 0.01	162.5	173.8	0.02	156.9	173.9	157.4	180.0
1990	183.2	188.8	< 0.01	172.4	183.7	< 0.01	163.8	180.8	164.5	187.0
1995	200.2	205.7	< 0.01	189.2	200.5	< 0.01	179.8	196.8	183.8	206.4
2000	204.6	210.2	< 0.01	192.4	203.7	< 0.01	173.0	190.0	173.2	195.8
2003	206.5	212.1	< 0.01	191.6	202.9	< 0.01	173.6	190.5	174.5	197.1

Table 1: Testing for the Number of Components, 1970-2003

two vs. three components. However, the p-values of this test are decreasing in the early 1970s and by 1976 the modified likelihood ratio test rejects a two component model at a level of 5%. In all subsequent years the hypothesis of two components is clearly rejected in favor of three components. This is also supported by the values of the model selection criteria AIC and BIC, which initially are in favor of a two component model, but over time switch toward the three component mixture model.

In Figure 2 we compare the fitted three-component density for 1976 and 2003 with a nonparametric density estimate with bandwidth $h_c(3)$. Such a comparison could also be used for a formal goodness of fit test for our mixture model, cf. e.g. Fan (1994). The nonparametric and our parametric estimate are quite close, thus, our model of the data seems appropriate. We furthermore give a Quantile-Quantile (QQ) plot of the data against the fitted mixture. Both plots clearly show that the three component mixture with equal variances adequately describes the data.

Mixture models are routinely used for discriminant analysis, see e.g. Fraley and Raftery (2002). In our analysis each observation can be assigned posterior probabilities which give the probability of the observation to belong to each of the components
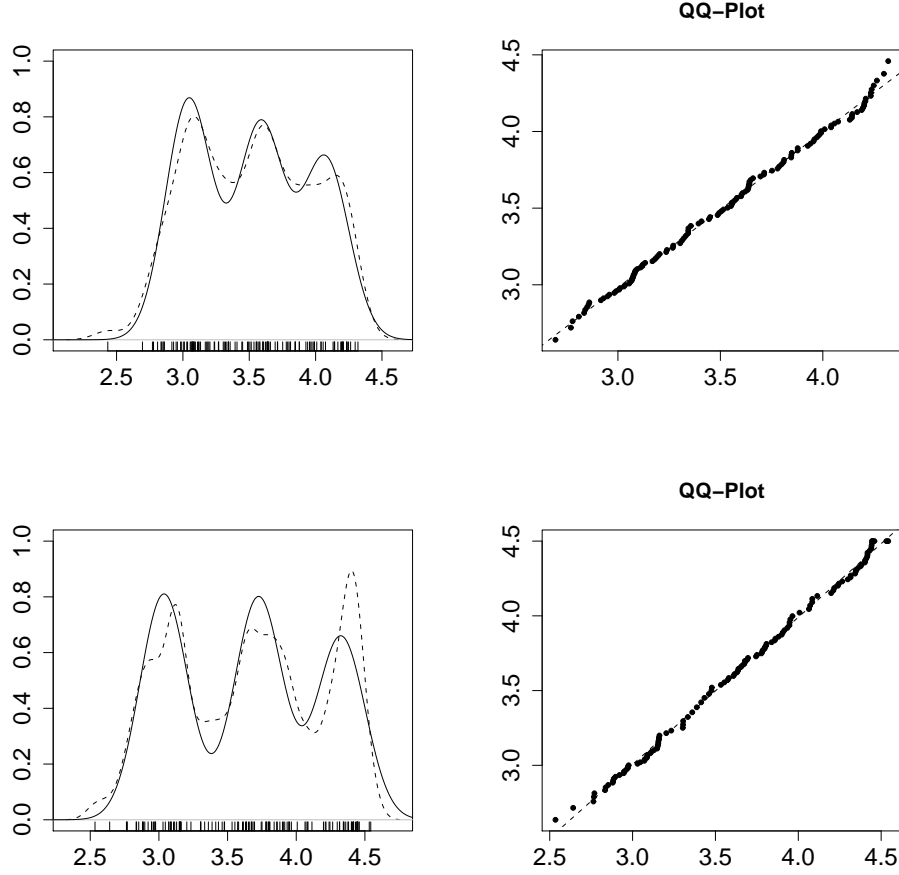
Figure 2: Up/left: Three-component mixture density with equal variances (solid line) and kernel density estimate based on $h_c(3)$ (dashed line) for the log-data (logarithm to the base 10) for 1976. Up/right: QQ-Plot of the log-data for 1976 against the quantiles of the normal mixture (three components, equal variances) together with least squares fit (dashed line). Down/left: Three-component mixture density with equal variances (solid line) and kernel density estimate based on $h_c(3)$ (dashed line) for the log-data (logarithm to the base 10) for 2003. Down/right: QQ-Plot of the log-data for 2003 against the quantiles of the normal mixture (three components, equal variances) together with least squares fit (dashed line).

in the mixture model. This yields three levels of income which we label poor, middle and rich, with indices $1, 2, 3$. The posterior probability of an observation y to belong to group j, $j = 1, 2$, is equal to

$$p_j(y) = \frac{\hat{\alpha}_j \phi(y; \hat{\mu}_j, \hat{\sigma})}{f(y; \hat{\alpha}_1, \hat{\alpha}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\sigma})},$$

and $p_3(y) = 1 - p_1(y) - p_2(y)$. Therefore, we do not merely assign an income level to each country, but rather a probability distribution, which makes transitions from one group to the other much more transparent. One may then assign and observation y to one of the components by using the maximum a-posterior estimate (MPE), which assigns the $j \in \{1, 2, 3\}$ to country i for which $p_j(y)$ is maximal. The posterior mean (PM) is the weighted average of the posterior probabilities, e.g. $PM(y) = p_1(y) + 2p_2(y) + 3p_3(y)$. One can also determine the thresholds $t_{j,j+1}$, $j = 1, 2$, for the values of log-GDP per capita at which the MPE changes between the state j and $j + 1$, by solving the equations $p_j(t_{j,j+1}) = p_{j+1}(t_{j,j+1})$, $j = 1, 2$, yielding the unique solutions

$$t_{j,j+1} = \frac{\hat{\mu}_j + \hat{\mu}_{j+1}}{2} + \hat{\sigma}^2 \frac{\log(\hat{\alpha}_j / \hat{\alpha}_{j+1})}{\hat{\mu}_{j+1} - \hat{\mu}_j}, \qquad j = 1, 2.$$

The percentage of countries ascribed to the first component dropped slightly over time from initially 37.9 percent in 1976 to 35.7 percent in 2003. In comparison, the second component slightly gained from 33.8 percent to 35.3 percent, leaving the third component weight more or less unaltered (28.4 percent in 1970 and 29 percent in 2003). We can observe in Figure 3 that the average GDP per capita of the first

component stagnated at a level slightly above $1100. The average GDP per capita of the second component increased from $3998 to $5504 which corresponds to a 37 percent increase between 1976 and 2003. The third component experienced an increase of average GDP per capita from $12335 to $21938 (increase of 77 percent). One should mention that up- and downward movements of countries affect these growth rates. The growth of the third component is slowed down by countries moving up from the second component. Regarding the second component there are positive and negative effects, in which the negative effects outweigh the positive effects, since only a few countries move from the third to the second component. In the first component there should be positive effects from countries coming from the second group, which however are counterbalanced by the poor overall growth record within this component.

With respect to movements between the three components, we find that seven countries move up from the first to the second component (China, Sri Lanka, India, Indonesia, Pakistan, Cameroon) and seven countries move up from the second to the third component (Korea, Taiwan, Equatorial Guinea, Cyprus, Malaysia, Mauritius, Chile). The speed of movement from one component to the other is quite rapid as one can see for selected countries in Figure 4.[3]

There are also some downward movements, five countries fall back from the second to the first component (Honduras, Cote d'Ivoire, Solomon Islands, Afghanistan, Iraq) and six countries fall back from the third to the second component (South Africa, Uruguay, Nicaragua, Argentina, Iran, Venezuela). However, their development from

---

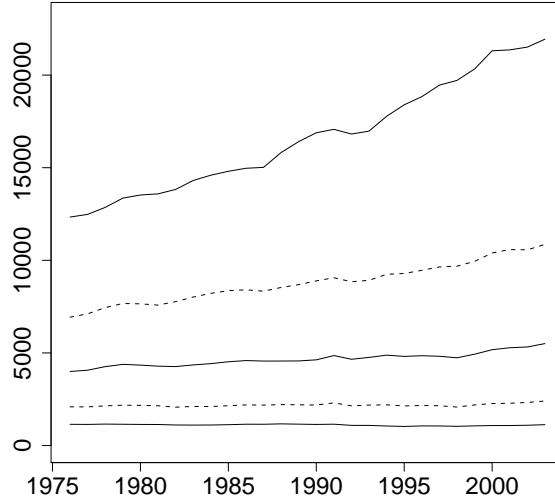[3] The picture is quite similar for the other movers.

12

Figure 3: Means of the distinct groups (solid lines). Income levels where the maximum a-posterior estimates switch from one group to the other (dashed lines).

1976 to 2003 seems to be heavily affected by external economic and political shocks and not so much by their long-run growth path.

The country specific posterior means help to explain the development of the cross-country distribution of GDP per capita from 1976 to 2003. The following general picture emerges: First, Sub-Saharan Africa accounts mostly for the first component which remains stagnant. Second, the emergence of the "transition"' component is mostly due to the growth "take off" in Asia and also to a relative decline of Latin America. Most Western countries belong firmly to the 3rd component displaying hardly any change in their posterior mean.
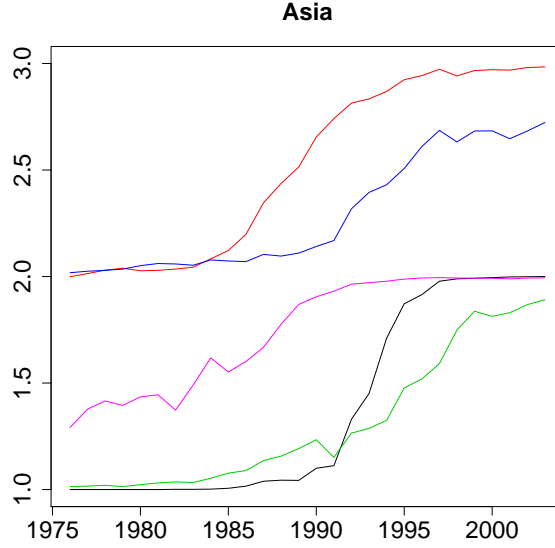
13

Figure 4: Posterior means of selected Asian countries: China (black), Republic of Korea (red), India (green), Malaysia (blue) and Indonesia (pink).

4   Discussion

In this paper we challenge the long standing twin peaks finding in the cross-country distribution of GPD per capita. We show that the number of peaks of a distribution depends on the scale (e.g. original or logarithmic) and argue that this makes the feature of twin peaks useless for economic interpretations. Instead we use finite mixtures to investigate the cross-country distribution of GDP per capita, since *a.* its number of components does not depend on the scale, *b.* components in the mixture arguably correspond better to income clubs in the distribution than its peaks, and *c.* finite mixtures allow for an accurate analysis of the intra-distributional dynamics by using posterior probability estimates.

14

We argue that components should take the place of twin-peaks in the economic growth literature. In contrast to twin-peaks, we find evidence for an emerging intermediate "transitional" component in the 1970s, resulting in a three-component distribution from 1976 onwards. The development of the three components confirms the theoretical predictions of the unified growth theory. The average GDP per capita of the first component does not change over the whole observation period and represents a regime of malthusian stagnation. The third component had by far the highest growth rates and represents a sustained growth regime. The second component represents countries on the move from one regime to the other, with constantly new countries taking off from the malthusian regime and others arriving at the sustained growth regime. Also the remarkable speed of the take off and the geographical clustering of the movers is consistent with the predictions of the theory.

In addition, our method might be a useful tool to classify countries into "poor", "medium" and "rich" groups. Due to its statistical nature, the approach would be less policy dependent than current approaches. The boundary points of income, separating the three groups, from our point of view currently somewhat arbitrarily obtained by the World Bank, could be replaced by the incomes where the maximum a-posterior estimate switches. For the year 2003 these are $2405 and $10859, respectively (PPP, base year 2000).

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov 6 F. Csaki (Eds.), *Second international symposium on information theory*, pp. 267-281.

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

Bianchi, M. (1997) Testing for Convergence: Evidence from Non-Parametric Multimodality Tests. *Journal of Applied Econometrics* **12**: 393–409.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2001) A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society Series B* **63**: 19–29.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2004) Testing for a finite mixture model with two components. *Journal of the Royal Statistical Society Series B* **66**: 95–115.

Chen, J. and Kalbfleisch,J. D. (2005) Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference* **129**: 93–107.

Dacunha-Castelle, D. and Gassiat, E. (1999) Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Annals of Statistics* **27**: 1178–1209

Fan, Y. (1994) Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory* **10**: 316–356.

Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**: 611–631.

Galor, O. (1996) Convergence? Inference from Theoretical Models. *Economic Journal* **106**: 1056–1069.

Galor, O. (2009) 2008 Lawrence R. Klein Lecture – Comparative Economic Development: Insights from Unified Growth Theory *Department of Economics, Brown University*.

Graham, B. S. and Temple, J. R. W. (2006) Rich nations, poor nations: how much can multiple equilibria explain? *Journal of Economic Growth* **11**: 5–41.

Heston, A.; Summers, R. and Aten, B. (2006) Penn World Tables Version 6.2. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.*

Jones, C. (1997) On the Evolution of the World Income Distribution. *Journal of Economic Perspectives* **11**: 19–36.

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

Paap, R. and Dieck, H. K. (1998) Distribution and mobility of wealth of nations. *European Economic Review* **42**: 1269–1293.

Pittau et al. (2010) Mixture Models, Convergence Clubs, and Polarization. *Review of Income and Wealth* forthcoming.

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

Quah, D. T. (1996) Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Economic Journal* **106**: 1045–1055.

Quah, D. T. (1997) Empirics for growth and distribution: Stratification, polarisation and convergence clubs. *Journal of Economic Growth* **2**: 27-59.

Sala-i-Martin, X. (1996) The Classic Approach to Convergence Analysis. *Economic Journal* **106**: 1019–1036.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, **6** : 461-464

Silverman, B. W. (1981) Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society Series B* **43**: 97–99.