The Millennium Peak in Club Convergence - A New Look at Distributional Changes in the Wealth of Nations

Melanie Krause*
University of Hamburg†

Draft Version, 1 September 2015

Abstract

The convergence debate of whether poorer countries are catching up with richer ones has recently focused on the concept of club convergence, hence convergence within groups of countries. Detecting club convergence in the distribution of countries' income per capita over time has, however, proved difficult. I propose a novel indicator that captures intradistributional changes in one number: With two clusters involved, changes in the critical bandwidth for unimodality reflect modes becoming more or less pronounced, which, respectively, is evidence for club convergence or de-clubbing. Significance of the change can be determined in a bootstrap procedure, while working with standardized densities removes the influence of time-varying variance. For a 123-country income per capita distribution, the new club convergence indicator shows that in the 1980s and 1990s, groups of poor and rich countries converged to two separate points. But this development peaked at the turn of the millennium and has since been followed by a de-clubbing movement, as some formerly poor countries are growing fast to catch up with the rich.

JEL Classification: C14, D3, O1, O4

Keywords: Club Convergence, Kernel Density Estimation, Multimodality, Bootstrap, Polarization

*Correspondence address: University of Hamburg, Department of Economics, Von-Melle-Park 5, 20146 Hamburg, Germany; E-mail: *melanie.krause@wiso.uni-hamburg.de*

†I am grateful for comments and suggestions to Michael Binder, Nicola Fuchs-Schündeln, Sebastian Kripfganz, Maxim Pinkovskiy and Nicholas Rohde as well as to seminar participants at Goethe University Frankfurt.

1 Introduction

Are poorer countries gradually catching up with their richer peers? This key question in international macroeconomics has spawned the huge growth literature on convergence.

The focus of the literature on convergence has in the past decades shifted from absolute convergence in GDP per capita across the world towards club convergence within groups of countries. This is motivated by two findings: When restricting the dataset to groups of similar countries, such as OECD countries, rather than looking at the global sample, studies are much more likely to report β-convergence (a negative relation between countries' initial GDP per capita level and their subsequent growth rates) and σ-convergence (a decreasing variance in countries' log GDP per capita).[1] Furthermore, absolute convergence across the whole distribution should manifest itself in a unimodal shape with ever-higher concentration around this mode. For some decades the distribution of GDP per capita has, however, exhibited a clearly multimodal shape, also called "Twin Peaks" by Quah (1996).
The term club convergence, as coined by Baumol (1986) and elaborated upon by Quah (1993, 1997), involves convergence of the GDP per capita levels only of countries in the same "club", of which there exist several. Theoretical models explaining the presence of these multiple steady states feature, among others, heterogeneity of technology, human capital and fertility across countries (Galor, 1996) or countries interacting with trading partners (Quah, 1996). But to what extent does the multimodality of the GDP per capita distribution really give evidence of such a club convergence process?

[1] For details on β-convergence see Barro (1991) and Mankiw et al. (1992), for σ-convergence Barro and Sala-i-Martin (2004). Overviews of the convergence literature are given by e.g. Temple (1999) and Islam (2003).

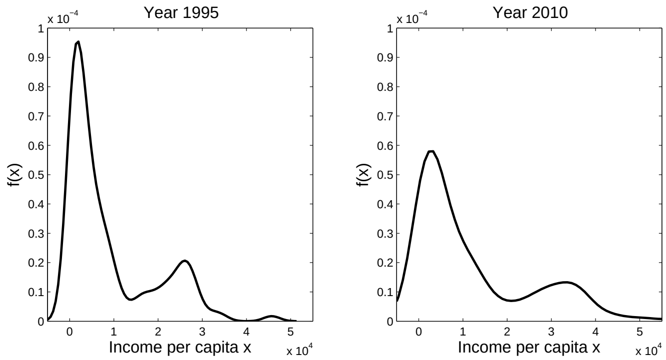

Figure 1: Kernel Density Estimation of the Absolute Income Per Capita Distribution Across the 123-Country Data Set in the Years 1995 and 2010
The graphs are kernel density estimates based on Gaussian kernel and Silverman's rule of thumb bandwidth.

Given the empirical relevance of club convergence, it is all the more unsatisfactory that this concept remains rather elusive from an econometric point of view. In the literature one cannot find an unambiguous formal definition for club convergence, nor a distribution-based test for it.[2] This is where this paper makes a contribution.

Consider the two plots of the income per capita distribution in 1995 and 2010 of a worldwide dataset comprising 123 countries (Figure 1). In both years the distribution clearly shows a high mode of poorer countries and a smaller one of rich countries. But this bimodal shape per se does not yet mean that club convergence has taken place between 1995 and 2010. In fact, if poorer and richer countries have converged towards separate points, these two modes must have become more pronounced over time. Now has this been the case? Visual inspection of intradistributional changes can be tricky and potentially misleading. The overall increase in mean income and in the distributional variance also complicates the direct comparison. And what conclusion on club convergence should the researcher draw if, say, one mode becomes

[2] There are panel data tests that can accommodate the club convergence hypothesis as cointegration between countries' income per capita time paths, such as the test by Hobijn and Franses (2000). However, these tests can be troubled by ex-ante assumptions for determining cluster size and membership, an issue that Canova (2004) addresses by working with Bayesian techniques. It would, nevertheless, be desirable to have a frequentist, nonparametric method for identifying club convergence to let the data speak for themselves when analyzing changes in the income per capita distribution.

more and the other one less pronounced?

The contribution of the paper is to propose a ready-to-use club convergence indicator that tracks over time how pronounced the multimodality of the distribution is. The new measure builds upon the literature of nonparametric multimodality tests proposed by Silverman (1981, 1983, 1986) and implemented by Bianchi (1997). These widely-used tests involve the calculation of the critical bandwidth for unimodality to test for the presence of multiple modes at a given point in time. Here the notion of the critical bandwidth is taken to the dynamic setting. The idea is that, when working with standardized densities, an increase in the critical bandwidth for unimodality over time indicates club convergence.
The proposed club convergence indicator has three appealing features: (i) It has an intuitive interpretation: If the two modes become more pronounced, more smoothing is necessary to achieve a unimodal distribution, hence the critical bandwidth for unimodality increases. (ii) It is statistically tractable: The significance of its change can be computed by a simple bootstrap procedure. (iii) It comes at a low extra cost: Researchers conducting the standard multimodality test already calculate the critical bandwidth, so tracing it over time is an easy and natural extension.
The new club convergence indicator provides new empirical insights into the evolution of the income per capita distribution of 123 countries from 1970 to 2011: In the 1980s and 1990s, groups of poor and rich countries converged to two separate points, but this club convergence movement peaked at the turn of the millennium. Since then, a significant de-clubbing movement can be observed, as modes are becoming less pronounced and some formerly poor countries are growing fast to catch up with the rich.

The rest of the paper is organized as follows: Section 2 briefly reviews the kernel density estimation literature and its application to the multimodality of the income per capita distribution. Section 3 contains the main contribution, namely constructing a club convergence indicator based on changes in the critical bandwidth. The significance of the changes is determined with the help of a bootstrap procedure. Section 4 lays out a comparison of the properties of the critical bandwidth to two polarization measures, which show interesting parallels to the club convergence concept. This leads to Section 5, which contains the empirical application showing the Millennium Peak in club convergence. Proofs and supplementary statistics have been relegated to the Appendix.

2 The Critical Bandwidth in a Static Setting

2.1 A Brief Review of Kernel Density Estimation

When researchers want to estimate the distribution of income per capita across countries without making any potentially limiting assumptions on its shape, they typically recur to the nonparametric technique of kernel density estimation. Being purely data-driven, this method allows to represent distributions that may be skewed, multimodal or have other characteristics that a parametric model cannot capture; for an introduction see Silverman (1986) as well as Bowman and Azzalini (1997). Assuming we observe n data points x_i $(i = 1, 2, ..., n)$, the kernel density estimate of density $f(x)$ is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big) \tag{1}$$

with kernel function K and bandwidth h. Heuristically, a density function of the form specified in the kernel is put around each of the observations and combined additively to the overall density function, using h as the smoothing factor. The widely-used Gaussian kernel[3]

$$K\Big(\frac{x - x_i}{h}\Big) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-x_i}{h})^2} \tag{2}$$

allows to write (1) as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2}. \tag{3}$$

The crucial choice in a kernel density estimation is the bandwidth h because, by regulating the amount of smoothing applied to the kernels around the data points,

[3] Other possible kernel functions include the Epanechnikov and the Triangular kernel. While the shape of the density is not crucially influenced by the kernel function, I will stick to the Gaussian kernel as it ensures the important analytical result by Silverman (1981) about the relation between the modality and the bandwidth, see Theorem 1 below.

it vitally determines the shape of the density and its modality.

Before going into more detail, let us impose the following standard regularity assumptions on the density, in line with Silverman (1983) and Mammen et al. (1992):[4]

Assumptions 1. *(a) f is a bounded density with bounded support on $[x_L; x_U]$.*

(b) f is twice continuously differentiable on $(x_L; x_U)$.

(c) On the boundaries of the density it holds: $f'(x_L+) > 0$ and $f'(x_U-) < 0$.

(d) The modality of f is the number of local maxima $\widetilde{x}$ where

$$\begin{cases} f'(\widetilde{x}) = 0, \\ f'(x) > 0 \quad \forall x \in \widetilde{X}_N \wedge x < \widetilde{x}, \\ f'(x) < 0 \quad \forall x \in \widetilde{X}_N \wedge x > \widetilde{x}, \end{cases}$$

with $\widetilde{X}_N$ denoting the neighborhood of the point $\widetilde{x}$. A j-modal density hence has j local maxima and $j-1$ local minima.

(e) For all points with $f'(x) = 0$, it holds that $f''(x) \neq 0$ and $f(x) > 0$.

(f) The first two moments of f exist and are finite.

Hence, the modality of a density is defined in terms of sign changes in its first derivative. To see the dependence of the density modality on the bandwidth used, let us consider an example. 300 observations are drawn from a Gaussian mixture with three components:

$$f(x) = \frac{2}{3} \cdot \phi(x, -2, 0.5^2) + \frac{1}{6} \cdot \phi(x, 0, 0.5^2) + \frac{1}{6} \cdot \phi(x, 2, 0.1^2), \tag{4}$$

where $\phi(x, \mu, \sigma^2)$ denotes the Gaussian distribution of x with mean μ and variance σ^2.

[4] Assumption (e) rules out certain turning points or "shoulders" in the density, which, even though they are typically estimated well in practice, can change the asymptotic properties, see Silverman (1983). Assumption (f) is an additional requirement which has to be imposed so that f can be standardized.

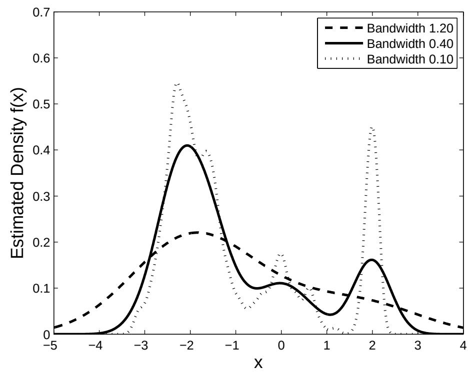

Figure 2: Kernel Density Estimation of (4) with Different Bandwidths

While the true underlying distribution has three modes, one can plot an estimated kernel density with *any* number m of modes ($1 \leq m \leq 300$) depending on the bandwidth used. Figure 2 shows three examples: A high bandwidth such as $h = 1.20$ induces so much smoothing that only the most pronounced mode remains. Gradually decreasing the bandwidth makes more modes appear, so that at $h = 0.40$ a trimodal distribution emerges. Further reducing the bandwidth reveals additional smaller features and spikes that can be considered as spurious modes, as in the seximodal distribution at $h = 0.10$.

Which bandwidth should one choose in practice when plotting a density? The vast literature on optimal bandwidth selection offers various techniques which strike a balance between bias and variance, ranging from Silverman's well-known rule of thumb (Silverman, 1986) to more sophisticated cross-validation methods (Jones et al., 1996) as well as adaptive bandwidth selection methods (see Cowell and Flachaire, 2015, for an overview). For our measure of club convergence, however, we are not concerned with the *optimal* bandwidth for appropriate representation but the *critical* bandwidth for m-modality.

2.2 The Critical Bandwidth for m-Modality and its Application to (Static) Multimodality Tests of the Income per Capita Distribution

The casual observation from Figure 2 that a lower bandwidth leads to the emergence of additional modes has been proved formally by Silverman (1981):

Theorem 1. *In a kernel density estimation of $f(x)$ with a Gaussian kernel, the number of modes is a right-continuous decreasing function of the bandwidth h.*

Proof. See Silverman (1981). □

He consequently defined the critical bandwidth for m-modality, CB^m, as the smallest bandwidth still producing an m-modal rather than $(m+1)-$modal density. For all bandwidths $h < CB^m$ the estimated density will have at least $m+1$ modes. CB^m can easily be computed by a binary search procedure.

Under the regularity conditions from Assumptions 1, one can derive the asymptotic properties of CB^m as the number of data points goes to infinity: Silverman (1983) and Mammen et al. (1992) show that for m-modal densities, CB^m converges to zero at the rate $n^{-\frac{1}{5}}$, while for densities with a higher modality, CB^m stays larger than a constant c_0:

Theorem 2. *Assume that the true density has j modes and that the regularity conditions from Assumptions 1 hold. For the critical bandwidth for m-modality, CB^m, it then holds:*

(a) If $m \geq j$, $CB^m \xrightarrow{p} 0$ at rate $n^{-\frac{1}{5}}$ as $n \to \infty$.

(b) If $m < j$, $P(CB^m > c_0) \to 1$ as $n \to \infty$.

Proof. See Silverman (1983) and Mammen et al. (1992) □

The key contribution contribution of this paper is to propose an indicator of club convergence based on changes in the critical bandwidth over time. This is a dynamic measure, which vitally sets itself apart from the current use of the critical bandwidth in a static setting. In the literature, a typical application of the critical bandwidth is for (static) multimodality tests: How many modes does the income per capita

distribution in a particular year have?
The null of m-modality is tested against the alternative hypothesis of more than m modes by calculating CB^m. A bootstrap resampling procedure helps to decide whether CB^m should be considered as too high for typical m-modal densities, which would lead to the conclusion that the density instead has at least $m + 1$ modes.[5] The (static) bootstrap multimodality test can be used successively with increasing m until one cannot reject the null anymore. It is not without critics: Mammen et al. (1992) were among the first to point out the conservatism of the test. There have since been some suggestions for ameliorations, for instance Hall and York's (2001) calibration adjustment.

The general concept has however remained unaffected and Silverman's (1981) bootstrap procedure based on CB^m remains one key method for testing for multimodality of a distribution at a given point in time. For instance, Bianchi (1997) applies it to income per capita data from 119 countries in 1970, 1980 and 1989 and finds that he cannot reject the null of unimodality in 1970, in contrast to the later years, where he finds evidence of bimodality.

His findings are qualitatively supported by alternative procedures to test for multimodality, in particular those involving mixture models. Paap and van Dijk (1998) model the income distribution of their 120-country sample from 1960 to 1989 with a mixture of a Weibull and a truncated Normal density. They identify these components with the help of the EM-algorithm by Dempster et al. (1977) and analyze the changes in the component variances. This is also the procedure employed by Pittau et al. (2010), who in an application with 102 workforce-weighted countries from 1960 to 2000 find three component densities.

Whether using the bootstrap test by Silverman (1981) or a mixture approach, these

[5] The bootstrap facilitates resampling from a density with a specified modality, CB^m, by adding a random component multiplied by a bandwidth. As implemented by Bianchi (1997), the sample variance adjustment by Efron and Tibshirani (1993) ensures that the smoothed bootstrap sample has the same variance as the original data. In the end, if the bootstrap's critical bandwidth $CB^{m\star}$ exceeds the original CB^m in relative terms less than test size α (e.g. 5%), the null of m-modality is rejected in favor of (at least) $(m + 1)$-modality.

tests for multimodality are notably static. They only allow to draw conclusions on the shape of the distribution of income per capita at a given point in time. The key argument in this paper is that in order to make inference about club convergence, one has to monitor the evolution of the distribution over time and in particular observe whether the modes become more pronounced. Visual comparisons of distributional features can however be tricky and potentially misleading, especially against the backdrop of an overall increase in variance. This can be resolved by looking at changes in the critical bandwidth, which, as I will argue now, capture the intradistributional changes underlying club convergence or de-clubbing.

3 An Indicator of Club Convergence based on Changes in the Critical Bandwidth

3.1 The Use of Standardized Densities

The key idea is to use changes in the critical bandwidth to measure how the shape of the distribution has changed: If the two modes of a bimodal distribution become more pronounced, the critical bandwidth for unimodality goes up as more smoothing must be applied to obtain a unimodal density.

A dynamic setting poses a challenge: The critical bandwidth based on raw data is sensitive to changes affecting the whole distribution, which is important in light of the well-known increase in worldwide variance. An indicator of club convergence should only reflect how pronounced the modes are relative to the other parts of the distribution and be invariant to changes in the overall distributional variance. This can be achieved by working with standardized densities that have the same shape as the original ones.[6]

Theorem 3. *Let $f(x)$ be a kernel density estimate with a Gaussian kernel, domain $[x_L; x_U]$ and bandwidth h_x. Standardize all n data points by subtracting the mean μ*

[6] While standardization removes the influence of both mean and variance, only the latter is crucial. A *ceteris paribus* increase in μ shifts the whole distribution, keeping the distance between the data points and hence CB^m unaffected. By contrast, an increase in σ clearly translates one-to-one into an increase in CB^m. A side benefit of centering around zero is that it facilitates the interpretation of the standardized data.

and dividing by the standard deviation σ, so that $y_i = \frac{x_i - \mu}{\sigma}$ for all $i = 1, ..., n$. The transformed density $f(y)$ then has the domain $\left[\frac{x_L - \mu}{\sigma}; \frac{x_U - \mu}{\sigma}\right]$ and it holds:

(a) $f(y) = \sigma f(x)$; hence the density values are scaled by σ.

(b) When employing the scaled bandwidth $h_y = \sigma^{-1} h_x$, the transformed density $f(y)$ has the same shape as the original $f(x)$.

Proof. See Appendix A1. □

Given that the original and standardized data have the same shape, it follows directly from Theorem 3:[7]

Corollary 1. *Standardization of the data does not affect the result of Silverman's (1981) multimodality test.*

3.2 Club Convergence and Changes in the Critical Bandwidth

In the following let us consider the setting of two country clusters that most empirical studies find for the worldwide income per capita distribution in the past decades. A cluster is a group of data points that can form a component in a mixture distribution. It is well-known that not every cluster will show itself in its own mode because for this it has to be sufficiently well-separated from the other cluster.[8] But no matter how well-pronounced the two modes are in the beginning, this paper argues that changes over time are what is crucial for the convergence debate. Only intradistributional changes can show the dynamics at work and potentially allow to make forecasts of the developments to come. In order to do that, let us now relate changes in the critical bandwidth for unimodality to club convergence:

[7] For the multimodality test in any given year it does not matter if the raw or standardized data are used: Data standardization scales both the critical bandwidth and the comparison critical bandwidths from the bootstrap resamples, leading to the same result as to how often one exceeds the other. However, to trace the critical bandwidth over time as the club convergence indicator does, the standardization is needed.

[8] Eisenberger (1964) derives necessary and sufficient conditions that the mean and variance of the two component densities of a Gaussian mixture have to satisfy in order to show themselves in two separate modes.

Definition 1. *Let $f(x)$ be a standardized income per capita density with at most two clusters that might or might not show themselves in two modes. The density is observed at two points in time, t=1,2 and the critical bandwidths for unimodality at $t = 1$ and $t = 2$ are calculated as CB_1 and CB_2. In this setting we say that we have*

- ***club convergence*** *if and only if $CB_2 > CB_1$.*
- ***de-clubbing*** *if and only if $CB_2 < CB_1$.*

Heuristically, when the two modes become more (less) pronounced, the critical bandwidth for unimodality,[9] CB, increases (decreases) because more (less) smoothing needs to be applied to make the bimodal shape turn into a unimodal one.

One caveat of this definition of club convergence is that it neglects the potential mobility of countries between clubs. But while it is advisable to check that the composition of the clubs is sufficiently stable over time, this seems to be a minor drawback in practice. In fact, there is strong empirical evidence for very limited mobility of countries between the clubs identified, see for instance Bianchi (1997) and Paap and van Dijk (1998).

On the other hand, a big advantage of Definition 1 is that by recurring to CB, it provides a club convergence indicator that captures the consequences of potentially complex intradistributional changes in just one number. In fact, club convergence can result from an increase in between-cluster separation, an increase of within-cluster concentration or a combination of both. As we will see, all of these developments will be reflected in an increase in CB:

Definition 2. *Starting with a density $f(x)$ consisting of two clusters, apply the following transformation to all observations $i = 1, ..., n_c$ in one cluster:*

$$y_i = x_i + a \tag{5}$$

- *If the transformation is applied to the poorer (richer) cluster and a is negative or*

[9]For notational convenience, I will henceforth write CB rather than CB^1 when it is clear that the critical bandwidth for unimodality is referred to.

- *if the transformation is applied to the richer (poorer) cluster and a is positive,*

density $f(y)$ is said to have an increased (decreased) ***between-cluster separation*** *with respect to $f(x)$.*

Corollary 2. *With CB_1 and CB_2 denoting, respectively, the critical bandwidth of the density before and after a ceteris paribus increase (decrease) in between-cluster separation, it holds: $CB_2 > (<) CB_1$.*

While there are a number of possibilities to alter the within-cluster concentration, let us focus on the λ-squeeze or λ-dispersion, which is common in the polarization literature (see Duclos et al., 2004). A λ-squeeze (λ-dispersion) leaves the mean of the cluster unchanged and decreases (increases) the within-cluster standard deviation by the factor λ:

Definition 3. *Starting with a density $f(x)$ consisting of two clusters, apply the following transformation to all observations $i = 1, ..., n_c$ in one or both clusters:*

$$y = \lambda x + (1 - \lambda)\mu_c, \tag{6}$$

where μ_c with $c = 1, 2$ denotes the mean of the respective cluster.

- *If $0 < \lambda < 1$, it is called a λ-squeeze, resulting in a higher* ***within-cluster concentration*** *in the respective cluster.*
- *If $\lambda > 1$, it is called a λ-dispersion, resulting in a lower* ***within-cluster concentration*** *in the respective cluster.*

Corollary 3. *With CB_1 and CB_2 denoting, respectively, the critical bandwidth of the density before and after a ceteris paribus λ-squeeze (λ-dispersion), it holds: $CB_2 > (<) CB_1$*

To understand the implications of Corollaries 2 and 3, let us consider the example of a data set containing 500 observations from the bimodal Gaussian mixture

$$f(x) = \frac{1}{2} \cdot \phi(x, 4, 1^2) + \frac{1}{2} \cdot \phi(x, 8, 0.5^2). \tag{7}$$

This exemplary income per capita distribution consists of a cluster of poorer countries with higher within-cluster variance (a more heterogeneous group) and a cluster of richer countries with a lower within-cluster variance. The critical bandwidth for unimodality of the standardized density is 0.7183. The second column of Table 1 shows some indicative values of this initial distribution, such as the difference between the modes and the within-cluster variances in relation to the overall variance. It will be instructive to compare these indicators, together with CB, for the transformed densities after conducting the ceteris paribus intradistributional changes (a) to (c). While these transformations concern the original data points, the changes in CB are calculated based on the standardized densities, filtering out the variance changes. Table 1 as well as Figure 3 show some characteristics of the raw and standardized densities before and after the transformations:

Value	Initial	(a)	(b)	(c)
Mode-distance (raw)	4	6	4	2.5
Mode-distance (stand.)	1.8669	1.9426	1.9494	1.8684
σ^2_{C1} (stand.)	0.2572	0.1238	0.0701	0.1649
σ^2_{C2} (stand.)	0.0498	0.0340	0.0543	0.1277
CB	**0.7183**	**0.8428**	**0.9072**	**0.7901**
p-value CB change	-	0.0024	0.0000	0.0778

Table 1: Descriptive Values of the Density (7) Before and After the Ceteris Paribus Changes in Cases (a) to (c)

The distance between the modes is calculated both for the raw and standardized density. σ^2_{C1} and σ^2_{C2} refer to the within-cluster variance of cluster 1 and 2 relative to the overall variance. It holds: $Var_{overall} = \frac{1}{1-n}[(n_{C1}-1)\sigma_1^2 + (n_{C2}-1)\sigma_2^2 + n_{C1}(\mu_1-\mu)^2 + n_{C2}(\mu_2-\mu)^2]$, where the two latter two terms refer to the between-cluster variance. CB denotes the critical bandwidth for unimodality. The p-value for insignificance of the change in CB is calculated based on the bootstrap procedure explained in Section 3.4 with 5000 replications.

(a) In the first case, the **between-cluster separation** is increased by adding $a = 2$ in (5) to the points in the rich cluster of the example distribution (7). The rich group of countries can be interpreted as moving away from the poorer ones. This is one typical case of club convergence, reflected by a substantial increase in CB of the standardized densities from 0.7183 to 0.8428, as Table 1, column (a), shows. The distance between the modes increases by construction, and, because the between-cluster variance has gone up, the relative within-cluster variance in both clusters decreases. This makes the modes more pronounced in the standardized densities in the upper right panel of Figure 3.

(b) A λ-squeeze with $\lambda = 0.5$ in (6) is applied to poor mode, **increasing its within-cluster concentration**. This is another typical club convergence example and CB clearly increases. Interestingly, the middle panel of Figure 3 shows that while the poorer mode becomes more pronounced, the richer one decreases in magnitude. This can be explained by the relative importance of the clusters: In kernel density estimation, the more concentrated a group of points, the higher their density values, at the expense of other parts of the density. Seeing one mode become more and the other one less pronounced clearly complicates visual inspection of club convergence, which shows the benefit of CB as an indicator.

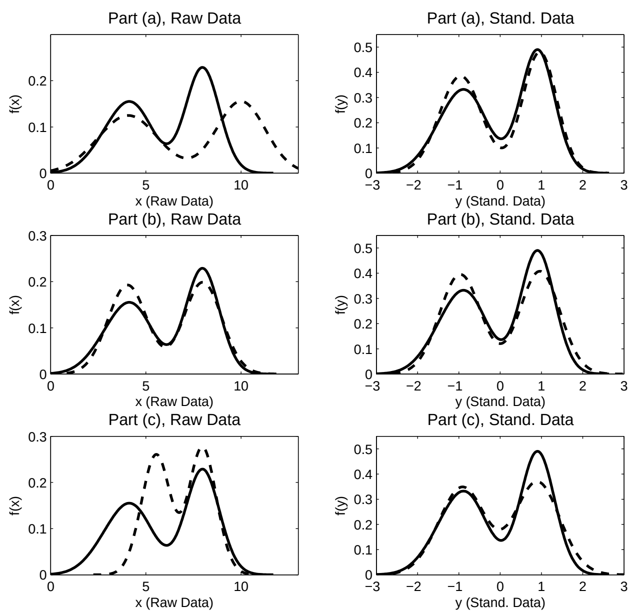

Figure 3: Raw and Standardized Densities Before (Solid Line) and After (Dashed Line) the Intradistributional Changes from Parts (a), (b) and (c)
The graphs are kernel density estimates based on Gaussian kernel and Silverman's rule of thumb bandwidth.

(c) What happens if a **decrease in between-cluster separation** ($a = 1.5$ for the poor cluster) is combined with an **increase in within-cluster concentration** (λ-squeeze with λ=0.5 in the poor cluster)? The visual inspection of the lower panel of Figure 3 is not very helpful in determining which of these two opposite effects dominates: In the raw density the clusters move together but with the poorer one becoming more concentrated. Filtering out the decrease of overall variance, the graph of the standardized densities displays the poorer mode as slightly more pronounced and the richer one as less so. This is also reflected in the within-cluster variances in column (d) of Table 1. Capturing the composite effect of these changes in one number, CB increases to 0.7901 and hence points to a club convergence movement, in which the increase in within-cluster concentration is larger than the decrease in between-cluster separation. Obviously, this hinges on the magnitude of the effects: Additional simulations show that when combining the same within-cluster concentration with a stronger decrease of between-cluster separation ($a = 2$ rather than $a = 1.5$), the latter effect dominates and CB decreases.

This example has illustrated how changes in the CB can capture in one number various intradistributional alterations. CB has the added advantage that it is readily available if the researcher anyways analyzes the multimodality of the density at a given point in time with Silverman's (1981) test: He or she merely has to work with the standardized densities and to track the test statistic over time. Incidentally, applying the static multimodality test with 5000 replications to the cases (a) to (c) would always reject the null unimodality can be rejected (p-value of 0) before and after the transformation. Hence, a static test can only conclude that the density is bimodal in each instance. By contrast, the changes in CB reflect vital intradistributional movements that can be interpreted as club convergence or de-clubbing.

3.3 Asymptotic Properties of the Club Convergence Indicator

In order to prove that the change in CB can be consistently estimated, one can extend the static results by Silverman (1983) and Mammen et al. (1992) presented in Section 2.2. When the number of data points goes to infinity, the change in CB based on the kernel density estimation is consistent and converges to a constant:

Theorem 4. *Assume that the density has up to two modes at times $t = 1, 2$ and that the regularity conditions from Assumptions 1 hold. Then,*

(a) the change in the critical bandwidth for unimodality, $CB_2 - CB_1$, based on the

kernel density estimates is a consistent estimator of the respective change in the true densities.

(b) $P(|CB_2 - CB_1| \geq c) \rightarrow 1$ *as* $n \rightarrow \infty$. *The constant* c *depends on the underlying distributional changes.*

Proof. See Appendix A2. □

3.4 Bootstrap Test for Significance of Changes in the Indicator

As with any measure that exhibits changes over time, in practical applications it is important to check the significance of an increase or decrease in CB. For the conclusion about club convergence and possible policy implications, it is vital that structural changes rather than random noise should be the cause of the change in CB. For this I suggest using a bootstrap procedure that incorporates longitudinal correlation in the spirit of Biewen (2002): For inference in the field of inequality, mobility and poverty, it is important to take into account the typically strong persistence of the data. The magnitude of intradistributional changes has to be examined against the backdrop of many countries keeping their place in the distribution.

A bootstrap can easily accomplish this by resampling from the same countries over time rather than taking random samples of the two distributions. This can be seen as a particular form of a block or cluster bootstrap, more precisely:

1. Start with two data sets $\mathbf{Y_1}$ and $\mathbf{Y_2}$ of the same countries $i = 1, ..., n$ observed at times $t = 1, 2$. Calculate the critical bandwidths for unimodality, CB_1 and CB_2, for both data sets.

2. Draw a bootstrap sample of size n of the numbers $1, ..., n$ with replacement, using them as indicators for the countries to be included in both $\mathbf{Y_1^\star}$ and $\mathbf{Y_2^\star}$.

3. For the samples $\mathbf{Y_1^\star}$ and $\mathbf{Y_2^\star}$, calculate the critical bandwidths for unimodality, $CB_1^\star$ and $CB_2^\star$.[10]

[10] Note that the setting in this bootstrap comparison of CB_1 to CB_2 is fundamentally different from Silverman's (1981) static bootstrap multimodality test. The latter draws smoothed bootstrap resamples from a density with a prespecified modality (marginally m-modal) in order to come to a conclusion about the m- vs. $(m+1)$-modality of the density. Here, we are concerned with the

4. Repeat Steps 2 and 3 a large number of times, each time storing the difference $CB_2^\star - CB_1^\star$. If more than $1 - \alpha$ (e.g. 95%) of all resampled differences $CB_2^\star - CB_1^\star$ are positive (negative), conclude that CB has increased (decreased) significantly.

This procedure allows to compare CB_2 to CB_1 in a purely data-driven way. Let us carry out the bootstrap procedure with 5000 replications for the distributional changes (a) to (c) considered above. The last row of Table 1 shows that the first two changes in CB are highly significant even at the 99% confidence level. In case (c), where effects work in opposite directions, the change in CB is smaller, but it is still significant at the 90% confidence level.

The bootstrap-based determination of significance in changes in CB will be important for the interpretation of the results in the empirical section. But in order to put this measure of club convergence and its properties into perspective, let us now compare CB to indices from the polarization literature.

4 Comparison of the Club Convergence Indicator to Polarization Measures

Conceptually different from the convergence literature, which analyzes changes in the income differences *between* countries over time, the polarization literature was developed to represent particular changes in the income distribution *within* a country. However, polarization measures are now also applied to the worldwide distribution of income between countries (see for example Pittau et al., 2010). The similarities in how these measures react to the intradistributional changes underlying club convergence make a direct comparison with respect to CB very appealing.

significance of the change in CB over time rather than the modality in a static sense. The reference value is a resampling from the two data sets, taking into account longitudinal correlation between countries.

4.1 Wolfson's (1994) Bipolarization Measure

Wolfson (1994) developed the notion of a country's bipolarization as the degree to which income is concentrated at both ends of the distribution rather than in the middle. His measure was constructed to capture the widely-discussed idea of "the disappearing middle class" and uses the formula

$$P_W = 2\frac{\mu}{m}\big(1 - 2L(0.5) - Gini\big), \tag{8}$$

where μ and m denote the mean and median of the distribution, $L(0.5)$ is the value of the Lorenz curve at the 50% income point (indicating the share of total income accruing to the poorer 50% of individuals) and $Gini$ is the Gini coefficient of inequality calculated as the expected mean difference between two incomes x_i and x_j,

$$Gini = \frac{E|x_i - x_j|}{2\mu}. \tag{9}$$

Wolfson (1994) shows that any symmetric, mean-preserving transfer on either side of the mean that leads to fewer inequality will increase bipolarization. An example is reproduced in Figure 4, where an initially uniform distribution ends up bimodal. Obviously, in such a situation CB would also increase considerably. If the situation described a relative distribution of income across countries (rather than within) and the transfer is interpreted as some countries growing faster than others, then we would be in a typical club convergence setting.

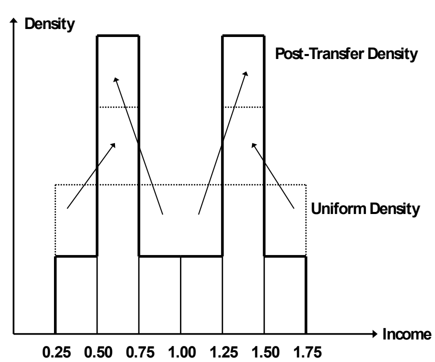

Figure 4: A Mean-Preserving Transfer Decreasing Inequality and Increasing Polarization based on Wolfson (1994)

4.2 The Polarization Measure by Esteban and Ray (1994) and Duclos et al. (2004)

Parallel to Wolfson (1994), Esteban and Ray (1994) as well as Duclos et al. (2004) developed and refined another polarization measure. They build on a political economy framework, with individuals identifying themselves with those of the same income and feeling alienation towards others. For continuous income distributions, the polarization formula is

$$P_{ER}^{\alpha}(f) = \int \int f(x)^{1+\alpha} f(y)|x-y|dxdy, \tag{10}$$

where $f(x)$ and $f(y)$ refer to the density values of income levels x and y, while the parameter $\alpha \in [0.25, 1]$ captures the degree of identification with people of the same income. The dependence on α can be seen as a drawback of this measure: Duclos et al. (2004) find that a ranking of countries according to their income polarization is sensitive to the value of α used.[11]

4.3 Properties of the Critical Bandwidth in Comparison to the Polarization Measures

Let us now derive several properties that CB fulfills and see to what extent they are shared by the polarization measures P_W and P_{ER}:

Theorem 5. *Let x be income per capita data whose density $f(x)$ has at most two clusters that might or might not show themselves in two modes. Calculate CB as the critical bandwidth for unimodality on standardized data $\frac{x-\mu}{\sigma}$, while P_W and P_{ER} are calculated based on raw data x as well as mean-standardized data $\frac{x}{\mu}$. Then it holds:*[12]

[11] Duclos et al. (2004) also point to an ambiguous association of polarization with multimodality in a general setting: The appearance of a third or fourth mode does not necessarily increase P_{ER} because it might decrease average income differences. This, however, is not relevant in the unimodal and bimodal setting in which we compare the polarization measures to CB.

[12] The polarization measures cannot be calculated based on standardized data: P_W includes the Gini coefficient and Lorenz curve, which do not allow negative income values. In the literature P_W is calculated either for raw data or mean-standardized data, which give the same result. P_{ER} is also often calculated on mean-standardized data. Duclos et al. (2004) derive the relation between their polarization measure for raw or mean-standardized data as $P_{ER}\left(\frac{x}{\mu}\right) = \mu^{\alpha-1} P_{ER}(x)$. Furthermore,

(a) ***General Structure:*** *A ceteris paribus increase (decrease) in between-cluster separation or within-cluster concentration leads to an increase (decrease) in CB, P_W and P_{ER}.*

(b) ***Scale Invariance:*** *If all incomes are scaled by a constant factor c as in $z = cx$,*

- *CB remains unchanged.*
- *P_W remains unchanged.*
- *$P_{ER}(\frac{x}{\mu})$ remains unchanged. $P_{ER}(x)$ is scaled: $P_{ER}(z) = c^{1-\alpha}P_{ER}(x)$.*

(c) ***Invariance to Absolute Income Changes:*** *When adding a constant amount a to all incomes as in $z = x + a$,*

- *CB remains unchanged.*
- *P_W changes to $P_W(z) = \frac{m_x}{m_x+a} \cdot P_W(x)$, where m_x is the median before the transformation.*
- *$P_{ER}(x)$ remains unchanged. $P_{ER}(\frac{x}{\mu})$ is scaled: $P_{ER}(\frac{z}{\mu_z}) = \left(\frac{\mu_x}{\mu_x+a}\right)^{1-\alpha} P_{ER}(\frac{x}{\mu_x})$.*

(d) ***Dispersion Invariance:*** *When applying a λ-squeeze ($0 < \lambda < 1$) or λ-dispersion ($\lambda > 1$) to all incomes so that $z = \lambda x + (1 - \lambda)\mu_x$,*

- *CB remains unchanged.*
- *P_W changes to $P_W(z) = \lambda \frac{m_x}{m_z} \cdot P_W(x)$.*
- *$P_{ER}(x)$ and $P_{ER}(\frac{x}{\mu})$ are scaled by the same factor to $P_{ER}(z) = \lambda^{1-\alpha}P_{ER}(x)$ and $P_{ER}(\frac{z}{\mu_z}) = \lambda^{1-\alpha}P_{ER}(\frac{x}{\mu_x})$.*

(e) ***Symmetry of the Polarization Measure ("Swapping Rich and Poor"):*** *When applying the transformation $z = x_L + x_U - x$ (with x_L and x_U denoting the infimum and the supremum of the income data),*

- *CB remains unchanged.*
- *P_W changes to $P_W(z) = \frac{m_x}{m_z} \cdot P_W(x)$, where m_x and m_z denote the median before and after the transformation.*

Esteban and Ray (2012) provide an axiomatic comparison of P_W and P_{ER}. As several of these axioms involve settings with more than two modes, it is not possible to use all of them here for a comparison with CB.

- *$P_{ER}(x)$ remains unchanged. $P_{ER}(\frac{x}{\mu})$ is scaled:*
  $P_{ER}(\frac{z}{\mu_z}) = \left(\frac{\mu_x}{x_L+x_U-\mu_x}\right)^{1-\alpha} P_{ER}(\frac{x}{\mu_x})$.

Proof. See Appendix A3. □

The conclusion from this comparison is that for CB only the shape of the distribution in terms of its modality matters, hence it is invariant to scalings, dispersions or uniform absolute increases. By contrast, "the size of the pie" to be distributed matters for both polarization measures. For instance, when all incomes increase by a positive additive component as in (c), P_W decreases because, heuristically, the poor now own a larger share of the total pie, even though the shape of the distribution has not changed.[13] In a similar way, P_W but not CB is sensitive to a transformation swapping the rich and the poor in an asymmetric distribution. Concerning P_{ER}, in Theorem 5 one again notes the dependence on the α-parameter, which shows up in the adjustment factors that many of the transformations entail. By contrast, CB is free of such parameters.[14]

5 Empirical Application: Club Convergence and De-Clubbing in the Wealth of Nations

In the empirical application I make use of the critical bandwidth for unimodality to gain new insights into the changing distribution of income per capita of countries around the world in the last decades. Have there been club convergence movements and which are the countries driving these developments?

[13]Chakravarty et al. (2007) have proposed a class of rather complex bipolarization measures that do fulfill the property of invariance to absolute income changes.

[14]Perhaps the polarization concept that is closest in spirit to my use of the critical bandwidth is proposed by Anderson's (2004), who, in a bimodal setting classifies intradistributional changes with the help of stochastic dominance conditions. However, the inclusion of the distributional covariances into the computation can make it challenging to implement when many time periods are involved. By contrast, CB captures the relevant intradistributional changes in one number and can easily be traced over many years.

5.1 The Data Set and Descriptive Statistics

From the Penn World Tables 8.0 (Feenstra et al., 2015) I take the *rgde* variable (real GDP at chained purchasing power parity 2005 US Dollars) and divide it by the country's population, *pop*, at the given point in time. This way the values of income per capita, x, for all countries and years are obtained.
For most countries the data is available from 1970 onwards and up to 2011, so these years mark the beginning and end of my sample period. For the analysis of distributional changes over time, it is vital that the data set is balanced so that the distribution consists of exactly the same countries over the years. Hence, I drop all countries for which the data is not available during the whole period; this eliminates in particular the ex-soviet republics which gained independence in 1990/1991. Also, it has become standard in the growth literature (Mankiw et al., 1992) to drop countries which are primarily oil producers or tiny states with a population below 300,000 because special economic conditions can be thought to apply there. In the end, my data set contains 123 countries from every region of the world. A list of all countries and their income values in 1970 and 2011 is contained in Appendix A4 (Tables 4 and 5).

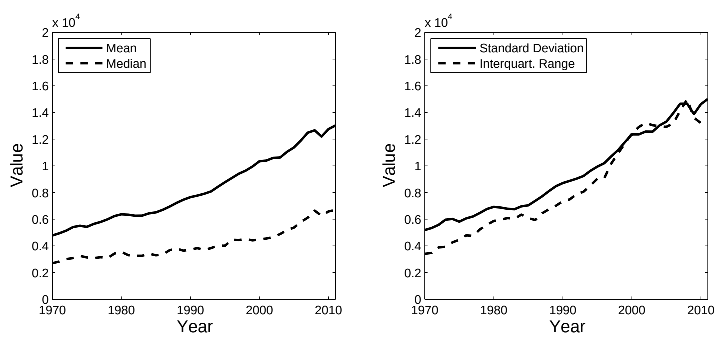

Figure 5: Descriptive Statistics for Income per Capita in the 123-Country Data Set

Figure 5 plots the evolution of some descriptive statistics of the data set over time. It is not surprising to see that both the mean and median of income per capita increase steadily throughout the last four decades, with the only short blip occurring in the aftermath of the global financial crisis in 2009. Mean income per capita was USD

4784 in 1970 and USD 13024 in 2011, which would equal an average yearly growth of $\left(\frac{\mu_{2011}}{\mu_{1970}}\right)^{\frac{1}{41}} - 1 \approx 2.47\%$. The fact that the median is below the mean reflects the positive skewness of the distribution. The second panel of Figure 5 displays measures of dispersion. In line with the evidence on σ-divergence in the literature, the standard deviation has increased steadily over time.[15] Looking at the individual countries in Tables 4 and 5, one can see that in 1970, income per capita ranges from USD 353 (Equatorial Guinea) to USD 23,659 (Switzerland), while 2011 it goes from USD 291 (Congo, Dem. Rep.) to USD 78,131 (Luxembourg). Against this backdrop, one might wonder whether the increase in standard deviation is driven by some outliers at the upper and lower tails of the distribution. This is however disproved by the second graph in Figure 5: The interquartile range, which measures the difference between the 25th and 75th percentile of the distribution and is robust to changes affecting only the tails, grows almost in sync with the standard deviation.[16] Hence, the increase in dispersion is driven by countries all over the distribution, which underpins my standardization procedure of subtracting the mean and dividing by the standard deviation.

Columns 5 and 6 of Tables 4 and 5 in Appendix A4 show the standardized income per capita values in 1970 and 2011 and can give an impression of countries' relative standing in the wealth of nations. The skewness of the distribution is reflected in the range of the values: The poorest countries are located at less than one standard deviation below the mean (Equatorial Guinea at -0.8539 in 1970, Congo, Dem. Rep. at -0.8485 in 2011) while the richest ones are more than three standard deviations above the mean (Switzerland at 3.6376 in 1970, Luxembourg at 4.3381 in 2011). Countries that grew at about the average rate of 2.47% over the sample period, like Peru or Greece, kept their standardized income values constant. But others improved their relative standing or fell behind, depending on their growth rates, which ranged from -2.92% (Liberia) to 8.27% (Equatorial Guinea). A regression of the growth rates on countries' initial income per capita levels, gives an insignificant

[15] σ-divergence denotes an increase in log variance of the distribution, for which an increase in variance is a necessary pre-condition.

[16] For comparison purposes with the standard deviation, the interquartile range is divided by a scaling factor of $2 \cdot 0.6745$, resulting from the interquartile range of the Gaussian distribution.

β-coefficient which is zero up to the seventh decimal, in line with the absence of unconditional β-convergence in the literature. But have the varying growth rates fostered the formation of convergence clubs? Let us now look at the shape of the distribution and its changes over time with the help of CB.

5.2 Results: Changes in the Distribution and the Critical Bandwidth

Kernel density estimations of the absolute income per capita values at four points in time are depicted in Figure 6. Obviously, the overall increase in variance can hide underlying intradistributional changes, making a look at the standardized densities in Figure 7 a bit more revealing. The big mode of relatively poor countries contains most of the mass throughout the sample period, but the smaller mode of richer countries close to two standard deviations above the mean evolves throughout the years, appearing more pronounced in 1985 than in 1970 and slightly less so at the end of the sample. But as was argued above, conclusions on club convergence based on visual inspection of changes in the distribution might be misleading, so let us now look at CB and its changes over time.

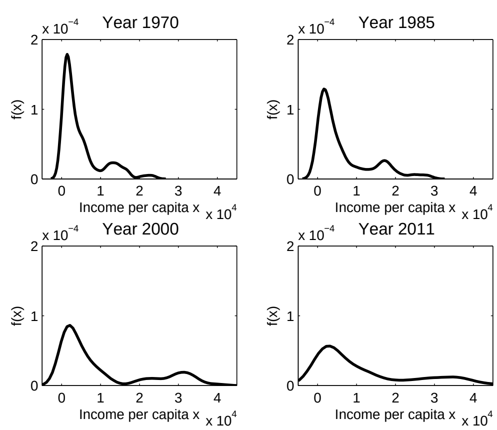

Figure 6: Kernel Density Estimation of the Absolute Income Per Capita Distribution Across the 123-Country Data Set in the Years 1970, 1985, 2000 and 2011
The graphs are kernel density estimates based on Gaussian kernel and Silverman's rule of thumb bandwidth.

When examining the modality of the distribution and its changes over time, we

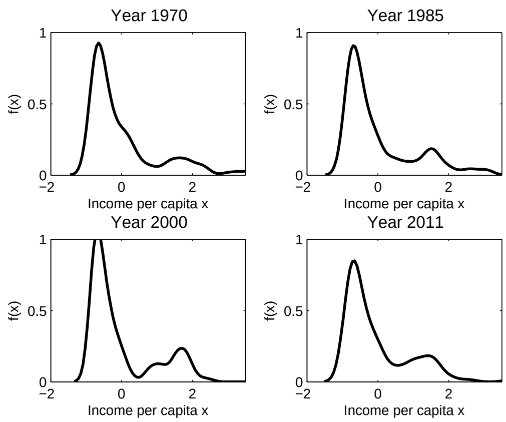

Figure 7: Kernel Density Estimation of the Standardized Income Per Capita Distribution Across the 123-Country Data Set in the Years 1970, 1985, 2000 and 2011
The graphs are kernel density estimates based on Gaussian kernel and Silverman's rule of thumb bandwidth.

have seen that CB has many benefits. However, one drawback of such a purely data-driven method in practice is its sensitivity to outliers. If a multimodality test were only based on changes in the first derivative of the density, a country such as Luxembourg, which in 2011 is four rather than two standard deviations above the mean, would constitute an individual mode. In the empirical multimodality literature, there are two simple approaches to deal with such isolated modes: Either eliminating outlier countries from the sample right away or including a threshold in the modality test that a density value has to exceed in order to classify as a mode. Here, I opt for the second possibility and find that for the standardized densities any threshold $f(x) = \tau$ with $\tau \in [0.02; 0.10]$ can be used to eliminate individual country modes while appropriately classifying larger clusters as modes.[17]

Silverman's (1981) bootstrap multimodality test clearly confirms that we are dealing with a density with up to two modes, once individual outlier modes are neglected: Conducting the (static) multimodality test for each year with 5000 bootstrap repli-

[17]Eliminating right-away the countries like Luxembourg, Macao and Switzerland that are outliers at some point in time would not affect the general results. Interestingly, the position of these countries varies as well: Switzerland forms an outlier mode in the 1970s but then falls slightly behind in relative terms and becomes absorbed into the rich mode, while Luxembourg, and most recently Macao, move away from the rich cluster to form individual modes.

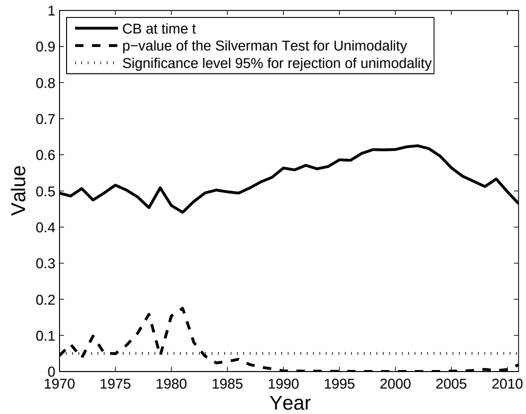

Figure 8: The Evolution of the Critical Bandwidth for Unimodality Over Time and p-Values of Silverman's (1981) Bootstrap Test for Unimodality with 5000 Replications

cations, one cannot reject the null of unimodality in the early years up to 1983. Figure 8 shows the p-values of the unimodality test (dashed line) together with the 5% horizontal line for significance at the 95% level. After 1984, unimodality is always rejected - and bimodality never - which is in line with the findings from other applications of multimodality tests in the literature, such as Bianchi (1997). But while such static tests can only conclude that there was unimodality until the 1980s and bimodality afterwards, we need a dynamic perspective to analyze the intradistributional changes over time.

Does the bimodal shape already imply that club convergence has taken place? The key insight from this paper is that this is not the case. The evolution of CB, the critical bandwidth for unimodality, plotted as a solid line in Figure 8, shows a rather nuanced picture: While the critical bandwidth varies around a constant level from 1970 to the middle of the 1980s, it exhibits a notable increase afterwards, but only until the turn of the millennium, when it peaks. The highest value of 0.6251 is reached in 2002. After that CB falls again until reaching levels of the 1970s and early 1980s. Hence, we observe temporary club convergence into two modes of rich and poor countries in the 1980s and 1990s, however, after the Millennium Peak, there is a tendency of de-clubbing with modes becoming again less clearly separated. These remarkable developments deserve a closer look.

Let us first assess the significance of changes in CB over time as proposed in Section 3.4. Figure 9 shows the results from bootstrap procedures with longitudinal correlation and 5000 replications: The p-values of a test of equality between CB in 1970 and later years (left panel) form a U-shape around the Millennium Peak. Hence, CB's increase in the 1980s/1990s makes it significantly higher in the late 1990s and early 2000s than at the beginning of the sample in 1970. The subsequent decrease means that in the late 2000s, CB is not significantly different from the 1970s anymore. The importance of this de-clubbing movement is confirmed in the right panel: From 2005 onwards, CB is already significantly lower than its 2002 peak value (equality test p-value of 0.01). Additional calculations with other reference years similarly elucidate the significance of the developments.

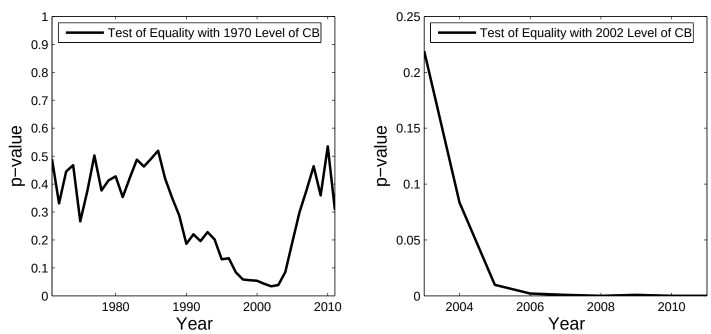

Figure 9: p-Values of Bootstrap Tests for Significance in the Change of CB since 1970 and 2002 (based on 5000 Replications)

5.3 Polarization Measures

Before analyzing in more detail the intradistributional changes leading to club convergence in the 1980s/1990s and de-clubbing in the 2000s, let us see to what extent these developments are also captured in the polarization and inequality measures. In particular, I calculate Wolfson's bipolarization index P_W from (8), the ER-polarization measure $P_{ER}(\alpha)$ from (10) (with $\alpha = 0.25$ and $\alpha = 1$, taking the two limits of admissible values to mark the range of the measure) as well as the Gini index (9) for (between-country) inequality. As was discussed in Section 4, these measures are calculated on mean-standardized data.

	CB	P_W	$P_{ER}(0.25)$	$P_{ER}(1)$	$Gini$
CB	1.0000	0.8109	0.8637	0.8558	0.7904
P_W	0.8109	1.0000	0.9772	0.8579	0.9580
$P_{ER}(0.25)$	0.8637	0.9772	1.0000	0.9035	0.9699
$P_{ER}(1)$	0.8558	0.8579	0.9035	1.0000	0.7865
$Gini$	0.7904	0.9580	0.9699	0.7865	1.0000

Table 2: Correlation Coefficients between CB and Polarization and Inequality Measures

Table 2 shows that over the sample period, CB has a high correlation with the polarization and inequality measures, even though these tend to have an even higher correlation among themselves. And when looking at the evolution of the measures over time in Figure 10 some interesting differences stand out - despite the Millennium Peak shared by all of them: In contrast to CB, Wolfson's bipolarization (left panel) starts to increase earlier (from the 1970s on) and more drastically so that even after the decline in the 2000s it stays well above its initial level. Hence, the income per capita distribution of countries was clearly more bipolarized in 2011 than in 1970. This evolution is similar to ER-Polarization with identification parameter $\alpha = 0.25$ (right panel), even if overall it exhibits less variation. On the other hand, ER-Polarization with $\alpha = 1$ behaves more similar to CB over the sample period, decreasing sizably after the Millennium Peak. The typically persistent Gini coefficient also shows a steady increase from 0.53 up to a Millennium Peak of 0.59, but afterwards remains at a rather high level of around 0.57.

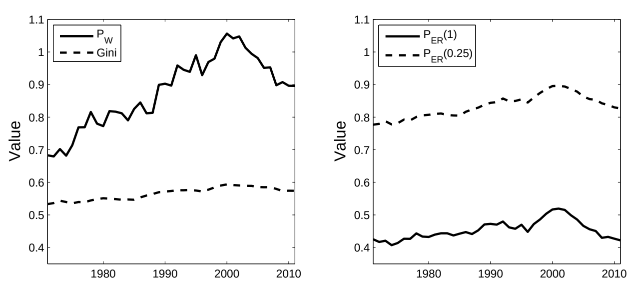

Figure 10: Measures of Polarization and Inequality Calculated for the 123-Country Data Set Over the Sample Period 1970-2011

.

A bootstrap test confirms the significance of these observed changes in the polarization and inequality measures:[18] For P_W, $P_{ER}(0.25)$ as well as $Gini$, one can reject equality with initial levels from the 1980s on, confirming our insight that the overall increase trumped the decrease after the Millennium Peak. Only $P_{ER}(1)$ shows a similar U-shape for the p-values as CB. One can conclude that, while the Millennium Peak appears in all polarization and inequality measures, only one shows a sufficiently strong decrease in the 2000s that mirrors the de-clubbing implied by CB.

5.4 Countries Driving the Club Convergence and De-Clubbing Movements

What has been going on beneath the surface to result in the club convergence and de-clubbing movements that the changes in CB indicate? Let us now turn to the country level and analyze which countries have been driving these developments. In order to reveal countries' club membership, we use the trough or antimode between the two modes in the kernel density plots from Figure 7 as the cut-off. In line with intradistributional movements, the trough fluctuates slightly between 0.5 and 1 standard deviation above the mean over the years. One has to keep in mind that with standardized data this whole analysis focuses on the relative rather than absolute per capita income. If a country grows at a moderate but below-average rate, it will fall behind its peers. The standardized data abstracts from the overall huge increase in income per capita: In 1970, the cut-off between the two modes lies at 1.00 standard deviation above the mean, which would correspond to an income per capita level of USD 9,950, while the cut-off in 2011 at 0.58 would correspond to an income per capita of USD 21,365.

The ensuing division of countries into the poor and rich club proves to be remarkably stable over time, confirming a key finding in the growth literature that comes with the club convergence interpretation. In fact, 109 out of the 123 countries in the data set stay in the same club for each of the 42 years from 1970 to 2011. Table 3 lists the countries in the poor and rich club at the end of the sample period, printing in bold the ones that changed clubs in between. The 28 countries in the rich

[18] The detailed results are available from the author upon request.

POOR CLUB IN 2011:

Albania, Argentina, Burundi, Benin, Burkina Faso, Bangladesh, Bulgaria, **Bahamas**, Belize, Bolivia, Brazil, Bhutan, Botswana, Central African Republic, Chile, China, Cote d'Ivoire, Cameroon, Congo (Rep.), Congo (Dem. Rep.), Colombia, Comoros, Cape Verde, Costa Rica, Djibouti, Dominican Republic, Ecuador, Egypt, Ethiopia, Fiji, Gabon, Ghana, Guinea, Gambia, Guinea-Bissau, Equatorial Guinea, Guatemala, Honduras, Hungary, Indonesia, India, Iran, Jamaica, Jordan, Kenya, Cambodia, Laos, Lebanon, Liberia, Sri Lanka, Lesotho, Morocco, Madagascar, Maldives, Mexico, Mali, Mongolia, Mozambique, Mauritania, Mauritius, Malawi, Malaysia, Namibia, Niger, Nepal, Pakistan, Panama, Peru, Philippines, Poland, Paraguay, Romania, Rwanda, Sudan, Senegal, Sierra Leone, El Salvador, Suriname, Swaziland, Syria, Chad, Togo, Thailand, **Trinidad & Tobago**, Tunisia, Turkey, Tanzania, Uganda, Uruguay, Vietnam, South Africa, Zambia, Zimbabwe

---

RICH CLUB IN 2011:

Australia, Austria, Belgium, Canada, Switzerland, **Cyprus**, Germany, Denmark, **Spain**, Finland, France, United Kingdom, **Greece**, **Hong Kong**, **Ireland**, Iceland, **Israel**, Italy, Japan, **Korea**, **Malta**, Netherlands, New Zealand, **Portugal**, **Singapore**, Sweden, **Taiwan**, United States

EXTREMELY RICH OUTLIER COUNTRIES IN 2011: Luxembourg, **Macao**

Table 3: Club Membership in 2011, Highlighting in Bold Countries Changing Clubs in 1970-2011

club are mainly highly-developed members of the OECD. But a closer look at the within-club heterogeneity as well as some insightful country trajectories over time shed more light on the club convergence and de-clubbing phenomena. Among the 14 "mobile" countries that changed clubs at least once during the 1970-2011 period, we find the Asian tigers (such as Korea and Taiwan) as well as countries from the European periphery (for instance Ireland, Spain, Cyprus) that typically managed the transition from the poor to the rich mode in the 1980s or early 1990s and have since become firmly established there. The club-changing countries also include Israel, which briefly dipped below the cut-off point into the poor group before returning to rich, or countries from the poor cluster, like Bahamas, which after some good years in the 1970s did not manage to stay in the rich group.

Figure 11 shows the trajectories of some selected countries in terms of standardized income per capita over the sample period and helps to illuminate the club convergence and de-clubbing movements. While these intradistributional developments can be complex in reality, the country-by-country examination reveals that they seem to be driven by a certain number of countries. In fact, most countries' relative income per capita positions are conspicuously stable, for example Germany or Kenya in the left panel of Figure 11. In contrast to that, countries near the threshold can vitally influence club convergence or de-clubbing movements: During the club convergence periods of the 1990s, there were relatively few countries near the cut-off threshold and both clusters became comparatively concentrated. One factor behind this may have been the fall of communism: Countries like Hungary or Poland, which had advanced towards the threshold from below, were temporarily thrown back into the poor mode. On the other hand, recent crossers into the rich mode, like Korea or Ireland, continued to grow and move away from the threshold. This development went on until the Millennium Peak in club convergence. After that, other driving forces led to clusters becoming less concentrated again and more countries approaching the threshold from both sides. There are a number of middle-income countries such as Poland, Turkey and Chile whose sustained growth has pushed them closer to the threshold in recent years. At the same time, we can see some countries from the rich cluster, for instance Greece, pictured in the right panel of Figure 11, perform badly and move closer to the threshold from above. Overall, this leads to the two modes becoming less pronounced, which is captured by the decrease in CB after the Millennium Peak.

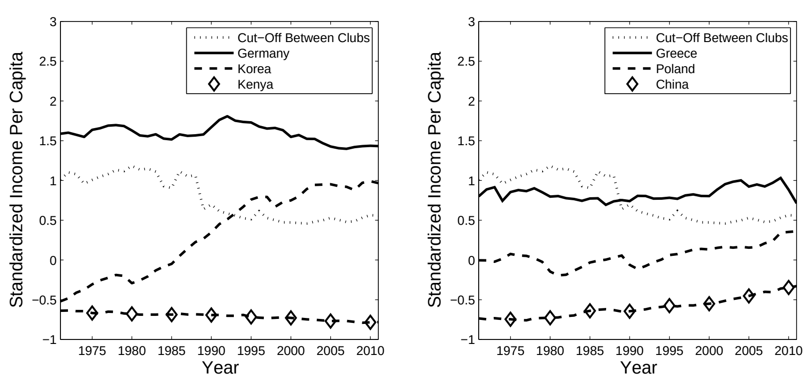

Figure 11: Trajectories of Selected Countries' Standardized Income per Capita Over Time

Which lessons can we draw from this analysis and which developments can we expect for the future? The idea of club convergence can be disheartening for poorer countries, as it would imply convergence to their own poor mode instead of aspiring to catch up with their richer peers. In this sense, it is positive to see that the de-clubbing movement has taken over from club convergence after the Millennium Peak. If middle-income countries keep up their above-average growth rates, they can be expected to pass the threshold into the rich cluster in the coming years.

Overall, if developments continue as they did in recent years, we should see more heterogeneity and less of a clear separation into a poor and rich mode, further decreasing the critical bandwidth for unimodality. Nevertheless, this does not mean that there will be (unconditional) β-convergence: Some poor countries like China in the left panel of Figure 11 have grown a lot in recent decades (even if it still has a long way to go to come close to the threshold), but there are others, particularly African countries, at the bottom of the distribution, which have stayed stuck there or fell behind even more. Further extrapolating into the future, there might be another club convergence movement into a very poor mode of countries left behind and a mode comprising the rich and growing middle-income countries. These are important insights, for instance for policymakers deciding on which countries to focus in their poverty reduction and development aid programs.
While the de-clubbing movement of recent years gives a message of potential mobility in the worldwide income per capita distribution, only the countries which have established an environment conducive to growth can reap the benefits.

6 Conclusion

In proposing a new measure of club convergence this paper has brought together three strands of the literature: Club convergence, kernel density estimation and polarization. Various intradistributional changes such as increases or decreases in between-cluster separation and within-cluster concentration are all summarized in just one number, which is easy to compute and interpret: An increase in the critical bandwidth for unimodality indicates club convergence, while a decrease corresponds to the opposite development of de-clubbing.

The application of my dynamic measure has lead to new empirical insights: With Silverman's (1981) static multimodality test one can only conclude that the income per capita distribution of 123 countries has been bimodal from the 1980s onwards but not how these two modes evolve over time. This is achieved by looking at changes in the critical bandwidth as a club convergence indicator: In the 1980s and 1990s, groups of poor and rich countries converged to two separate points, but this club convergence movement went only until the turn of the millennium. Since this so-called Millennium Peak in Club Convergence, a significant de-clubbing movement can be observed. As some formerly poor countries are growing fast to catch up with the rich and heterogeneity within clusters has increased, the modes are becoming less pronounced.
The comparison between the club convergence indicator and the polarization indices has elucidated some parallels, however, the former is the only measure to focus exclusively on changes in the shape of the distribution. This might explain, why, despite their overall high correlation, the polarization measures fail to show the strong de-clubbing movement since the turn of the millennium.
It is now up to further research - and the political debate - how individual countries can grow in order to achieve a prosperous position in the distribution.

References

Anderson, G. (2004). Towards an Empirical Analysis of Polarization. *Journal of Econometrics 122*, 1–26. DOI:10.1016/j.jeconom.2003.10.017.

Barro, R. J. (1991). Economic Growth in a Cross-Section of Countries. *Quarterly Journal of Economics 106*, 407–443. DOI:10.2307/2937943.

Barro, R. J. and X. Sala-i-Martin (2004). *Economic Growth*. McGraw Hill.

Baumol, W. (1986). Productivity Growth, Convergence, and Welfare. *American Economic Review 76*, 1072–1085.

Bianchi, M. (1997). Testing for Convergence: Evidence from Non-Parametric Multimodality Tests. *Journal of Applied Econometrics 12*, 393–409. DOI:10.1002/(SICI)1099-1255(199707)12:4<393::AID-JAE447>3.0.CO;2-J.

Biewen, M. (2002). Bootstrap Inference for Inequality, Mobility and Poverty Measurement. *Journal of Econometrics 108*, 317–342. DOI:10.1016/S0304-4076(01)00138-5.

Bowman, A. W. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford University Press.

Canova, F. (2004). Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach. *International Economic Review 45*, 49–77. DOI:10.1111/j.1468-2354.2004.00117.x.

Chakravarty, S., A. Majumder, and S. Roy (2007). A Treatment of Absolute Indices of Polarization. *Japanese Economic Review 58*, 273–292. DOI:10.1111/j.1468-5876.2007.00332.x.

Cowell, F. and E. Flachaire (2015). Statistical methods for distributional analysis. In A. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, pp. 359–466. North Holland.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum Likelihood from Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society B 69*, 1–38.

Duclos, J.-Y., J. Esteban, and D. Ray (2004). Polarization: Concepts, Measurement, Estimation. *Econometrica 72*, 1737–1772. DOI:10.1111/j.1468-0262.2004.00552.x.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

Eisenberger, I. (1964). Genesis of Bimodal Distributions. *Technometrics 6*, 357–363. DOI:10.1080/00401706.1964.10490199.

Esteban, J. and D. Ray (1994). On the Measurement of Polarization. *Econometrica 62*, 819–852. DOI:10.2307/2951734.

Esteban, J. and D. Ray (2012). Comparing Polarization Measures. In M. Garfinkel and S. Skarperdas (Eds.), *Oxford Handbook of Economics of Peace and Conflict*, pp. 127–151. Oxford University Press.

Feenstra, R. C., R. Inklaar, and M. P. Timmer (2015). The Next Generation of the Penn World Table. *American Economic Review 105*, 3150–3182. Available for download at www.ggdc.net/pwt.

Galor, O. (1996). Convergence? Inferences from Theoretical Models. *The Economic Journal 106*, 1056–1069. DOI:10.2307/2235378.

Hall, P. and M. York (2001). On the Calibration of Silverman's Test for Multimodality. *Statistica Sinica 11*, 515–536.

Hobijn, B. and P. Franses (2000). Asymptotically Perfect and Relative Convergence of Productivity. *Journal of Applied Econometrics 15*, 59–81. DOI:10.1002/(SICI)1099-1255(200001/02)15:1<59::AID-JAE544>3.0.CO;2-1.

Islam, N. (2003). What have we Learnt from the Convergence Debate? *Journal of Economic Surveys 17*, 309–360. DOI:10.1111/1467-6419.00197.

Jones, M., J. Marron, and S. Sheather (1996). A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association 91*, 401–407. DOI:10.1080/01621459.1996.10476701.

Mammen, E., J. Marron, and N. Fisher (1992). Some Asymptotics for Multimodality Tests based on Kernel Density Estimation. *Probability Theory and Related Fields 91*, 115–132. DOI:10.1007/BF01194493.

Mankiw, N. G., D. Romer, and D. Weil (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics 107*, 407–437. DOI:10.2307/2118477.

Paap, R. and H. K. van Dijk (1998). Distribution and Mobility of Wealth of Nations. *European Economic Review 42*, 1269–1293. DOI:10.1007/978-0-387-72796-7_5.

Pittau, M. G., R. Zelli, and P. Johnson (2010). Mixture Models, Convergence Clubs, and Polarization. *Review of Income and Wealth 56*, 102–122. DOI:10.1111/j.1475-4991.2009.00365.x.

Quah, D. (1993). Empirical Cross-Section Dynamics in Economic Growth. *European Economic Review 37*, 426–434. DOI:10.1016/0014-2921(93)90031-5.

Quah, D. (1996). Twin Peaks: Growth Convergence in Models of Distribution Dynamics. *Economic Journal 106*, 1045–1055. DOI:10.2307/2235377.

Quah, D. (1997). Empirics for Growth and Distribution: Stratification, Polarization and Convergence Clubs. *Journal of Economic Growth 2*, 27–59. DOI:10.1023/A:1009781613339.

Silverman, B. (1981). Using Kernel Density Estimation to Investigate Multimodality. *Journal of the Royal Statistical Society, Series B (Methodological) 43*, 97–99.

Silverman, B. (1983). Some Properties of a Test for Multimodality Based on Kernel Density Estimates. In J. Kingman and G. Reuter (Eds.), *Probability, Statistics and Analysis*, pp. 248–259. Cambridge University Press.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature 37*, 112–156. DOI:10.1257/jel.37.1.112.

Wolfson, M. (1994). When Inequalities Diverge. *American Economic Review, Papers and Proceedings 84*, 401–421.

7 Appendix

7.1 Appendix A1: Proof of Theorem 3 on the Shape of a Standardized Density

Proof. In order to prove this theorem, one can make use of the following result for the density of a transformed variable:

$$y = u(x) \quad \Longrightarrow \quad f_y(y) = \left|\frac{\partial(u^{-1}(y))}{\partial y}\right| \cdot f_x(u^{-1}(y)) \tag{11}$$

In our case the original density is the Gaussian kernel density estimate

$$f_x(x) = \frac{1}{nh_x}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-x_i}{h_x}\right)^2}. \tag{12}$$

The standardization is applied to the n data points x_i so that $y_i = \frac{x_i - \mu}{\sigma}$, for all $i = 1, ..., n$. In order to obtain a kernel density estimate of the standardized data points, the domain points x also have to be standardized: $y = \frac{x-\mu}{\sigma}$. This gives the transformation

$$y = u(x) = \frac{x - x_i}{\sigma} + y_i. \tag{13}$$

The two statements of the proposition can now be proved:

(a) Plugging (13) into (11) gives

$$\begin{aligned} f_y(y) &= \frac{\partial(\sigma(y - y_i) + x_i)}{\partial y} \cdot \frac{1}{nh_x} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_i}{h_x}\right)^2} \qquad (14) \\ &= \sigma \frac{1}{nh_x} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-x_i}{h_x}\right)^2} = \sigma f_x(x) \qquad (15) \end{aligned}$$

(b) (11) can be rewritten involving the bandwidth h_y, when it is defined as $h_y = \sigma^{-1}h_x$, as well as by substituting $x - x_i$ from (13):

$$f_y(y) = \frac{1}{n\frac{h_x}{\sigma}} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sigma(y-y_i)+x_i-x_i}{h_x}\right)^2} = \frac{1}{nh_y} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-y_i}{h_y}\right)^2} \tag{16}$$

Hence, when estimating $f_y(y)$ using $h_y = \sigma^{-1}h_x$ as the bandwidth, the direct correspondence of density values from (15) holds and ensures that the shape of the density is unchanged.

□

7.2 Appendix A2: Proof of Theorem 4 on the Consistency of the Change in the Critical Bandwidth

Proof. Statement (a) of Theorem 4 concerns the consistency of the estimated difference in $CB_2 - CB_1$ as $n \to \infty$. From Theorem 2, proved in Silverman (1983) and Mammen et al. (1992), it follows that at times $t = 1, 2$, CB_1 and CB_2 based on the kernel density are consistent estimates for $n \to \infty$. As a linear combination, their difference $CB_2 - CB_1$ is a consistent estimate as well.

Concerning the lower bound of the difference in statement (b) of Theorem 4, let us distinguish the different cases of unimodality ($j = 1$) and bimodality ($j = 2$) at times $t = 1, 2$:

- In the trivial but special case of $j = 1$ both at times $t = 1, 2$, by Theorem 2, part (a), it holds that both CB_1 and CB_2 converge to zero at the same rate of $n^{-\frac{1}{5}}$. Hence,

$$CB_2 - CB_1 \xrightarrow{p} 0 \text{ and } c = 0 \text{ in } P(|CB_2 - CB_1| \geq c) \to 1. \tag{17}$$

  An intradistributional change that keeps the density unimodal will not affect CB when $n \to \infty$.

- Assume now that $j = 1$ at $t = 1$ and $j = 2$ at $t = 2$. While CB_1 asymptotically converges to zero, CB_2 is bounded from below by a constant $c_0 > 0$ in Theorem 2. Hence, the difference $CB_2 - CB_1$ also converges to a positive constant and one could pick $c = c_0 > 0$ as a lower bound in $P(|CB_2 - CB_1| \geq c) \to 1$. So the change from a unimodal to a bimodal shape is reflected in the fact that the difference in CB is strictly positive, also asymptotically. By the same token, in the converse case of $j = 2$ at $t = 1$ and $j = 1$ at $t = 2$, $CB_2 - CB_1$ converges to a negative constant but by working with the absolute value, $c = c_0 > 0$ as a lower bound in $P(|CB_2 - CB_1| \geq c) \to 1$ still holds.

- Finally, consider $j = 2$ both at times $t = 1, 2$ but an intradistributional change that, without loss of generality, leads to $CB_2 > CB_1$. From Theorem 2, part (b), both CB_1 and CB_2 are asymptotically bounded by constants $c_0^{(1)}$ and $c_0^{(2)}$:

$$P(CB_1 > c_0^{(1)}) \to 1 \text{ and } P(CB_2 > c_0^{(2)}) \to 1 \tag{18}$$

  However, due to the intradistributional change leading to $CB_2 > CB_1$, one can pick $c_0^{(1)} < c_0^{(2)}$ and define $c = c_0^{(2)} - c_0^{(1)}$ in

$$P(|CB_2 - CB_1| \geq c) \to 1. \tag{19}$$

  This completes the proof of Theorem 4.

□

7.3 Appendix A3: Proof of Theorem 5 on the Comparison of the Critical Bandwidth to Polarization

Proof. (a) In a bimodal setting, the reaction of CB, P_W and P_{ER} to a ceteris paribus increase (decrease) in between-cluster separation or within-cluster concentration follows directly from the definition of the measures, see Corollaries 2 and 3 in Section 3 for CB as well as Esteban and Ray (2012) for P_W and P_{ER} .

(b) Scale Invariance:

- The standardized values on which CB is calculated are unaffected by a multiplicative factor c in $z = cx$: $\frac{z-\mu_z}{\sigma_z} = \frac{cx-c\mu_x}{c\sigma_x} = \frac{x-\mu_x}{\sigma_x}$.
- $$P_W(z) = 2\frac{\mu_z}{m_z}\big(1 - 2L_z(0.5) - Gini_z\big) \tag{20}$$
  with $z = cx$ can be reduced to $P_W(x)$ by making use of $L_z = L_x$, $Gini_z = Gini_x$, $\mu_z = c\mu_x$ as well as $m_z = cm_x$.[19]
- In
  $$P_{ER}^{\alpha}(z, w) = \int\int f(z)^{1+\alpha} f(w)|z - w|dzdw, \tag{21}$$
  one can make use of
  $$z = u(x) \quad \Longrightarrow \quad f_z(z) = \left|\frac{\partial(u^{-1}(z))}{\partial z}\right| \cdot f_x(u^{-1}(z)). \tag{22}$$
  to obtain the density $f(cx)$ (and equivalently $f(cy)$). With $z = cx = u(x)$, $f(z) = \frac{1}{c}f(x)$ as well as $dz = c \cdot dx$, one gets
  $$P_{ER}^{\alpha}(z, w) = \int\int \Big(\frac{1}{c}f(x)\Big)^{1+\alpha}\frac{1}{c}f(y)|cx - cy|c^2dxdy = c^{1-\alpha}P_{ER}(x, y).$$
  This is the homogeneity of degree zero property Duclos et al. (2004) point out: Mean-standardizing ($c = \frac{1}{\mu}$) the data scales P_{ER} by $\mu^{\alpha-1}$. It can directly be used to prove the second part of the statement, namely the scale-invariance for polarization based on mean-standardized data:
  $$\begin{aligned} P_{ER}\Big(\frac{z}{\mu_z}\Big) &= \mu_z^{\alpha-1}P_{ER}(z) = (c\mu_x)^{\alpha-1}c^{1-\alpha}P_{ER}(x) \\ &= \mu_x^{\alpha-1}\mu_x^{1-\alpha}P_{ER}\Big(\frac{x}{\mu_x}\Big) = P_{ER}\Big(\frac{x}{\mu_x}\Big) \end{aligned} \tag{23}$$

(c) Invariance to Absolute Income Changes:

[19] This is also the reason why P_W is the same both on raw and mean-standardized data: Mean-standardization can be seen as a scaling by a constant $c = \frac{1}{\mu}$.

- The standardized values on which CB is calculated are unaffected by an additive constant a in $z = x + a$: $\frac{z-\mu_z}{\sigma_z} = \frac{x+a-(\mu_x+a)}{\sigma_x} = \frac{x-\mu_x}{\sigma_x}$.
- Similar to part (a), $P_W(z)$ with $z = x + a$ can be expressed in terms of $P_W(x)$, using $\mu_z = \mu_x + a$, $m_z = m_x + a$ and $Gini_z = \frac{\mu_x}{\mu_x+a} Gini_x$. With

$$L_x(0.5) = \frac{\sum_{j=1}^{0.5n} x_j}{n\mu_x} \tag{24}$$

one can write $L_z(0.5)$ in terms of $L_x(0.5)$:

$$L_z(0.5) = \frac{\sum_{j=1}^{0.5n} z_j}{\sum_{k=1}^{n} z_k} = \frac{\sum_{j=1}^{0.5n} x_j + 0.5 \cdot n \cdot a}{n(\mu_x + a)} = \frac{L_x(0.5) \cdot n \cdot \mu_x + 0.5 \cdot n \cdot a}{n(\mu_x + a)} \tag{25}$$

Hence, $P_W(z)$ can be expressed as

$$P_W(z) = 2\frac{\mu_x + a}{m_x + a}\Big(1 - 2\frac{L_x(0.5) \cdot \mu_x + 0.5 \cdot a}{\mu_x + a} - \frac{\mu_x}{\mu_x + a} Gini_x\Big), \tag{26}$$

which after some algebra simplifies to $P_W(z) = \frac{m_x}{m_x+a} \cdot P_W(x)$. Increasing all incomes by a positive amount will thus decrease bipolarization.
- Following the same steps as in part (b) with the transformation $z = x + a$, one can directly see that the densities and differences involved do not change, hence $P_{ER}(z) = P_{ER}(x)$. Together with part (b) this implies for the mean-standardized data:

$$P_{ER}\Big(\frac{z}{\mu_z}\Big) = \mu_z^{\alpha-1} P_{ER}(z) = \Big(\frac{1}{\mu_x + a}\Big)^{1-\alpha} P_{ER}(x) = \Big(\frac{\mu_x}{\mu_x + a}\Big)^{1-\alpha} P_{ER}\Big(\frac{x}{\mu_x}\Big)$$

(d) Dispersion Invariance:

- The standardized values on which CB is calculated are unaffected by a λ-squeeze or λ-dispersion $z = \lambda x + (1-\lambda)\mu_x$: $\frac{z-\mu_z}{\sigma_z} = \frac{\lambda x+(1-\lambda)\mu_x-\mu_x}{\lambda\sigma_x} = \frac{x-\mu_x}{\sigma_x}$.
- Similar to (b) and (c), with $z = \lambda x + (1-\lambda)\mu_x$ it holds that $\mu_z = \mu_x$, while $m_z \neq m_x$ unless the distribution is symmetric. Plugging $Gini_z = \lambda Gini_x$ as well as

$$L_z(0.5) = \frac{\sum_{j=1}^{0.5n} z_j}{\sum_{k=1}^{n} z_k} = \frac{\sum_{j=1}^{0.5n} (\lambda x_j + (1-\lambda)\mu_x)}{n\mu_x} = \lambda L_x(0.5) + 0.5(1-\lambda) \tag{27}$$

into the formula for $P_W(z)$ yields $P_W(z) = \lambda \frac{m_x}{m_z} \cdot P_W(x)$.

- One can follow the same steps as in part (b) to show that $P_{ER}(z) = \lambda^{1-\alpha} P_{ER}(x)$ because the transformation $z = \lambda x + (1-\lambda)\mu_x$ induces the same changes to the density and differences as $z = cx$ with $c = \lambda$. For the mean-standardized data it holds:

$$P_{ER}\Big(\frac{z}{\mu_z}\Big) = \mu_z^{\alpha-1} P_{ER}(z) = \mu_x^{\alpha-1}\lambda^{1-\alpha} P_{ER}(x) = \lambda^{1-\alpha} P_{ER}\Big(\frac{x}{\mu_x}\Big) \tag{28}$$

(e) Symmetry of the Polarization Measure - Swapping Rich and Poor:

- The transformation equals a reflection of the distribution along the vertical line at $\frac{x_L + x_u}{2}$. The density values of two points of the distribution are swapped, $f(z) = f(x_U + x_L - x)$, which also holds for the standardized densities. The modality of the distribution - and hence CB - are unaffected by this symmetric reflection.
- For P_W one can proceed analogously to parts (b) to (d) and derive $\mu_z = x_L + x_U - \mu_x$, $Gini_z = \frac{\mu_x}{\mu_z} Gini_x$ as well as

$$L_z(0.5) = \frac{-L_x(0.5)\cdot\mu_x + 0.5(x_L + x_U)}{x_L + x_U - \mu_x}. \tag{29}$$

  Substitution into the formula yields $P_W(z) = \frac{m_x}{m_z} P_W(x)$.
- Following the same steps as in part (b) with the transformation $z = x_L + x_U - x$, one can see the densities and differences involved do not change and $P_{ER}(z) = P_{ER}(x)$. For the mean-standardized data:

$$\begin{aligned} P_{ER}\Big(\frac{z}{\mu_z}\Big) &= \mu_z^{\alpha-1} P_{ER}(z) = (x_L + x_U - \mu_x)^{\alpha-1} P_{ER}(x) \\ &= \Big(\frac{\mu_x}{x_L + x_U - \mu_x}\Big)^{1-\alpha} P_{ER}\Big(\frac{x}{\mu_x}\Big). \end{aligned} \tag{30}$$

□

7.4 Appendix A4: Overview of Countries in the Data Set

Country code	Country	Absolute Income per Cap.		Stand. Income Per Cap.		Growth per Year
		1970	2011	1970	2011	(1970-2011)
ALB	Albania	3747.64	7364.71	-0.1998	-0.3771	0.0166
ARG	Argentina	2950.94	14507.62	-0.3534	0.0988	0.0396
AUS	Australia	16750.87	38499.27	2.3062	1.6974	0.0205
AUT	Austria	12406.66	37282.53	1.4690	1.6164	0.0272
BDI	Burundi	558.01	490.14	-0.8145	-0.8352	-0.0032
BEL	Belgium	14252.76	35446.27	1.8248	1.4940	0.0225
BEN	Benin	1153.63	1231.88	-0.6997	-0.7857	0.0016
BFA	Burkina Faso	503.75	1051.52	-0.8250	-0.7978	0.0181
BGD	Bangladesh	1364.96	1554.21	-0.6590	-0.7643	0.0032
BGR	Bulgaria	3023.84	12906.67	-0.3393	-0.7800	0.0360
BHS	Bahamas	12045.06	19366.61	1.3993	0.4226	0.0117
BLZ	Belize	3510.06	7366.64	-0.2456	-0.3770	0.0182
BOL	Bolivia	1629.12	4166.78	-0.6081	-0.5902	0.0232
BRA	Brazil	3116.37	9294.53	-0.3215	-0.2485	0.0270
BTN	Bhutan	1084.23	4607.02	-0.7131	-0.5609	0.0359
BWA	Botswana	706.120	11810.75	-0.7860	-0.0809	0.0711
CAF	Central African Republic	1032.93	617.29	-0.7230	-0.8267	-0.0125
CAN	Canada	16064.66	35344.87	2.1740	1.4872	0.0194
CHE	Switzerland	23658.73	44823.64	3.6376	2.1188	0.0157
CHL	Chile	6336.88	15243.33	0.2992	0.1479	0.0216
CHN	China	966.92	8068.60	-0.7357	-0.3302	0.0531
CIV	Cote d'Ivoire	2363.19	1371.83	-0.4666	-0.7764	-0.0132
CMR	Cameroon	1233.60	1857.53	-0.6843	-0.7441	0.0100
COD	Congo (Dem. Rep.)	836.55	290.63	-0.7609	-0.8485	-0.0255
COG	Congo (Rep.)	1270.84	2426.87	-0.6772	-0.7061	0.0159
COL	Colombia	4025.07	8407.92	-0.1463	-0.3076	0.0181
COM	Comoros	1166.41	921.28	-0.6973	-0.8064	-0.0057
CPV	Cape Verde	965.64	4125.81	-0.7360	-0.5929	0.0361
CRI	Costa Rica	5446.86	10123.36	0.1277	-0.1933	0.0152
CYP	Cyprus	5797.19	28183.25	0.1952	1.0101	0.0393
DEU	Germany	12944.22	34519.98	1.5726	1.4323	0.0242
DJI	Djibouti	5402.75	2391.99	0.1192	-0.7084	-0.0197
DNK	Denmark	16978.34	35641.17	2.3501	1.5070	0.0183
DOM	Dominican Republic	2705.56	8726.60	-0.4600	-0.2864	0.0290
ECU	Ecuador	2533.23	6828.09	-0.4339	-0.4129	0.0245
EGY	Egypt	905.40	4836.37	-0.7476	-0.5456	0.0417
ESP	Spain	9549.38	28740.77	0.9183	1.0472	0.0272
ETH	Ethiopia	556.73	782.71	-0.8148	-0.8157	0.0083
FIN	Finland	13099.40	33747.33	1.6025	1.3808	0.0233
FJI	Fiji	2951.07	4644.74	-0.3533	-0.5583	0.0111
FRA	France	14512.68	31437.94	1.8749	1.2269	0.0190
GAB	Gabon	5351.55	12402.88	0.1093	-0.0414	0.0207
GBR	United Kingdom	13004.91	32259.81	1.5843	1.2817	0.0224
GHA	Ghana	2114.91	2522.37	-0.5145	-0.6998	0.0043
GIN	Guinea	1590.00	958.320	-0.6156	-0.8040	-0.0123
GMB	Gambia	1266.55	1236.29	-0.6780	-0.7854	-0.0006
GNB	Guinea-Bissau	1230.59	906.67	-0.6849	-0.8074	-0.0074
GNQ	Equatorial Guinea	353.50	9175.83	-0.8539	-0.2564	0.0827
GRC	Greece	8588.25	23698.65	0.7331	0.7112	0.0251
GTM	Guatemala	2889.37	4235.90	-0.3652	-0.5856	0.0094
HKG	Hong Kong	6777.86	38568.79	0.3842	1.7021	0.0433
HND	Honduras	2108.75	2919.84	-0.5157	-0.6733	0.0080
HUN	Hungary	4940.08	18852.01	0.0300	0.3883	0.0332
IDN	Indonesia	825.20	4339.49	-0.763	-0.5787	0.0413
IND	India	1222.28	3601.68	-0.6865	-0.6278	0.0267
IRL	Ireland	8125.97	36704.62	0.6440	1.5778	0.0375
IRN	Iran	3028.48	11818.47	-0.3384	-0.0803	0.0338
ISL	Iceland	14466.64	31921.62	1.8660	1.2591	0.0195
ISR	Israel	11729.06	25081.19	1.3384	0.8034	0.0187
ITA	Italy	11089.52	29089.05	1.2152	1.0704	0.0238
JAM	Jamaica	5474.39	5078.14	0.1330	-0.5295	-0.0018
JOR	Jordan	2702.27	5092.50	-0.4013	-0.5285	0.0156
JPN	Japan	11451.39	30427.21	1.2849	1.1596	0.0241

Table 4: First Part of the Countries in the Data Set

For the beginning and end of the sample, 1970 and 2011, both absolute and standardized values of income per capita are given. Absolute values are expressed in PPP 2005 USD; standardization is carried out by subtraction of the mean and division by the standard deviation. The growth rate is the average yearly growth rate for the country based on the 1970 and 2011 absolute values.

Country code	Country	Absolute Income per Cap. 1970	Absolute Income per Cap. 2011	Stand. Income Per Cap. 1970	Stand. Income Per Cap. 2011	Growth per Year (1970-2011)
KEN	Kenya	1474.72	1297.57	-0.6379	-0.7814	-0.0031
KHM	Cambodia	1298.75	2347.91	-0.6718	-0.7114	0.0145
KOR	Korea Rep	1903.57	27522.30	-0.5552	0.9660	0.0673
LAO	Laos	654.52	2623.87	-0.7959	-0.6930	0.0344
LBN	Lebanon	5189.38	13158.62	0.0780	0.8900	0.0230
LBR	Liberia	1596.03	474.47	-0.6145	-0.8362	-0.0292
LKA	Sri Lanka	2560.14	4701.08	-0.4287	-0.5546	0.0149
LSO	Lesotho	536.82	1487.82	-0.8186	-0.7687	0.0252
LUX	Luxembourg	22242.02	78130.59	3.3645	4.3381	0.0311
MAC	Macao	5327.00	69471.51	0.1046	3.7611	0.0646
MAR	Morocco	1914.60	3647.45	-0.5531	-0.6248	0.0158
MDG	Madagascar	1327.29	759.41	-0.6663	-0.8172	-0.0135
MDV	Maldives	1108.23	10343.66	-0.7085	-0.1786	0.0560
MEX	Mexico	6929.52	12709.82	0.4134	-0.0210	0.0149
MLI	Mali	452.41	941.06	-0.8349	-0.8051	0.0180
MLT	Malta	6220.10	23993.08	0.2767	0.7309	0.0335
MNG	Mongolia	958.07	5219.47	-0.7374	-0.5200	0.0422
MOZ	Mozambique	408.34	817.70	-0.8434	-0.8133	0.0171
MRT	Mauritania	1665.12	2615.75	-0.6012	-0.6935	0.0111
MUS	Mauritius	3806.06	9645.06	-0.1886	-0.2252	0.0229
MWI	Malawi	774.91	802.26	-0.7727	-0.8144	0.0008
MYS	Malaysia	2743.98	13468.81	-0.3932	0.0296	0.0396
NAM	Namibia	4142.39	5146.14	-0.1237	-0.5249	0.0053
NER	Niger	1030.42	522.560	-0.7235	-0.8330	-0.0164
NLD	Netherlands	14861.05	38054.85	1.9420	1.6678	0.0232
NPL	Nepal	754.17	1185.38	-0.7767	-0.7888	0.0111
NZL	New Zealand	14157.92	26666.53	1.8065	0.9090	0.0156
PAK	Pakistan	1453.33	2472.89	-0.6420	-0.7031	0.0130
PAN	Panama	4630.29	12154.75	-0.0297	-0.0579	0.0238
PER	Peru	3357.04	8923.98	-0.2751	-0.2732	0.0241
PHL	Philippines	2076.39	3521.06	-0.5219	-0.6332	0.0130
POL	Poland	4616.73	18430.43	-0.0323	0.3602	0.0343
PRT	Portugal	6807.18	22289.90	0.3898	0.6174	0.0294
PRY	Paraguay	1815.40	4351.30	-0.5722	-0.5779	0.0216
ROU	Romania	2526.24	13574.31	-0.4352	0.0366	0.0419
RWA	Rwanda	971.20	1201.50	-0.7349	-0.7878	0.0052
SDN	Sudan	1010.46	2373.99	-0.7273	-0.7096	0.0211
SEN	Senegal	1633.78	1411.72	-0.6072	-0.7738	-0.0036
SGP	Singapore	5262.33	51643.66	0.0921	2.5732	0.0573
SLE	Sierra Leone	1182.68	867.03	-0.6941	-0.8101	-0.0075
SLV	El Salvador	816.88	1116.53	-0.7646	-0.7934	0.0077
SUR	Suriname	4156.31	6699.65	-0.1210	-0.4214	0.0117
SWE	Sweden	16515.69	36100.79	2.2609	1.5376	0.0193
SWZ	Swaziland	1504.57	4239.25	-0.6321	-0.5854	0.0256
SYR	Syria	2743.49	3919.02	-0.3933	-0.6067	0.0087
TCD	Chad	1123.32	1851.12	-0.7056	-0.7445	0.0123
TGO	Togo	1082.39	946.69	-0.7135	-0.8047	-0.0033
THA	Thailand	1982.10	8491.04	-0.5401	-0.3021	0.0361
TTO	Trinidad & Tobago	9203.12	20196.31	0.8516	0.4779	0.0194
TUN	Tunisia	2200.01	6632.04	-0.4981	-0.4259	0.0273
TUR	Turkey	5732.40	14437.29	0.1827	0.0941	0.0228
TWN	Taiwan	3770.13	28413.56	-0.1955	1.0254	0.0505
TZA	Tanzania	1287.40	1269.39	-0.6740	-0.7832	-0.0003
UGA	Uganda	985.33	1187.03	-0.7322	-0.7887	0.0046
URY	Uruguay	7049.31	12625.06	0.4365	-0.0266	0.0143
USA	United States	20494.50	42646.21	3.0277	1.9737	0.0180
VNM	Vietnam	700.06	3447.77	-0.7872	-0.6381	0.0397
ZAF	South Africa	5312.42	8457.45	0.1018	-0.3043	0.0114
ZMB	Zambia	3873.56	2051.71	-0.1755	-0.7311	-0.0154
ZWE	Zimbabwe	2128.35	4347.79	-0.5119	-0.5781	0.0176
	Sample Mean	4784.40	13024.30	0	0	0.0272
	Sample Stand. Deviation	5188.72	15008.01	1	1	0.2449

Table 5: Second Part of the Countries in the Data Set

For the beginning and end of the sample, 1970 and 2011, both absolute and standardized values of income per capita are given. Absolute values are expressed in PPP 2005 USD; standardization is carried out by subtraction of the mean and division by the standard deviation. The growth rate is the average yearly growth rate for the country based on the 1970 and 2011 absolute values.