Q-convergence with interquartile ranges

INSTITUTE OF POLICY AND PLANNING SCIENCES

Discussion Paper Series

NO. 923

Q-Convergence with Interquartile Ranges

by

Sung Jin Kang, Myoungjae Lee

May 2001

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Q-Convergence with Interquartile Ranges

Sung Jin Kang and Myoungjae Lee
Institute of Policy and Planning Sciences
University of Tsukuba,
Ibaraki, 305-8573, Japan
Email: sungjin@sk.tsukuba.ac.jp; mjlee@sk.tsukuba.ac.jp

May 8, 2001

Abstract

We introduce a new convergence concept "Q-convergence" which defines convergence in national incomes as a shrinking interquartile range (IQR) of the national income distribution. Compared with the other convergence definitions in the literature, Q-convergence has advantages of taking into account clustering as well as dispersion of the income distribution; also, IQR is insensitive to outliers and equivariant to log-transformation, leading to robust statistical inference and easier reconciliation of the empirical findings using level and log data. A panel data is analyzed to find that the absolute income gap between the poor and rich countries has increased in terms of IQR, but the widening gap is rather small and insignificant when compared with the income increase of the poor countries.

Key words: convergence of income, quantile, panel data

JEL classification: C14, C33, O40, O50.

1 Introduction.

Convergence, *poor countries catching up with rich countries in national income or GDP per worker*, has been one of the controversial issues in economic growth. There are a number of formal definitions for the verbal description. One is "β-convergence": with y_0 and y_1 denoting the base and comparison period income or GDP per worker respectively, the growth rate $(y_1 - y_0)/y_0$ is a decreasing function of y_0. Despite the popularity, this definition has the critical weakness of involving only one country over time whereas the verbal description of convergence involves *many countries in two groups (rich and poor) over time.*

Another definition is "σ-convergence": with SD denoting standard deviation, the SD of the comparison period is smaller than the SD of the base period, i.e., $SD(y_1) < SD(y_0)$. By looking at SD at each period, this definition does involve many countries over time, differently from β-convergence. A shortcoming of σ-convergence is however that dispersion does not capture the dichotomy of rich and poor countries. It is known that β-convergence does not necessarily imply σ-convergence, while β-divergence does imply σ-divergence. See Barro and Sala-i-Martin (1995), Durlauf and Quah (1999), and the references therein for more on β- and σ-convergence.

Yet another strand of convergence definition is based upon "classification" or "clustering", motivated by the dichotomy of the poor and the rich: there is only one cluster at period 1 while there were two at period 0 (e.g. "twin-peaks" in Quah (1996a, 1996b)). This approach also involves many countries over time, but has a shortcoming that the notion of dispersion is not necessarily reflected in clusters. Bianchi (1997) compares the number of modes of the GDP distribution at period 0 and 1 to conclude convergence if the number of modes has declined ("modal convergence"), which however does not necessarily imply a declining SD.

σ-convergence and clustering approaches share the two disadvantages. One is lack of invariance (or equivariance) to monotonic transformations; here, we have log-transformation in mind which is popular in practice. The other is that there are many ways to measure/define dispersion and clusters.

Facing these problems in defining convergence, we propose a new definition "Q-convergence": in a nutshell, our proposal is examine the interquartile range (IQR) in the base and comparison period to conclude convergence if IQR has decreased. Compared with the other convergence concepts in the literature, this approach has the following three advantages, among which the first is trivial and the last two will be examined closely in the next section:

First, IQR is a well-known measure of dispersion as SD is, but differently from SD, IQR is not sensitive to outliers.

Second, the lower quartile (i.e., 25% quantile) and the upper quartile (i.e., 75% quantile) reflect the centers of two clusters (under one condition), much as two modes do in a bimodal distribution.

Third, the quartiles are equivariant to non-decreasing transformations, and thus inference based on IQR is more easily interpretable when y is subject to the popular log-transformation.

Thus, using IQR avoids the aforementioned problems in the convergence literature.

All convergence definitions have the "marginal convergence" version when only y is used, and a "conditional convergence" version when y is used with some variables, say x, controlled for which are relevant for the steady state of the economy. Controlling for x can be done by estimating the conditional distributions of $y_0|x$ and $y_1|x$ nonparametrically, or by using parametric regression residuals instead of y. In the literature, the term "absolute convergence" has been used instead of marginal convergence. But marginal convergence seems more fitting vis-á-vis conditional convergence, and we will use the term "absolute convergence" for a different purpose later in this paper.

In Section 2, the advantages of IQR are examined closely and our test statistics are presented formally along with their asymptotic distributions. In Section 3, we present our empirical analysis. Here, first, β-, σ- and modal convergence in the literature are applied, and then Q-convergence; these four convergence concepts are then compared. Finally, in Section 4, conclusions are drawn.

2 Interquartile Range for Convergence.

In this section, we examine the last two advantages of Q-convergence in detail. Since the third - more easily interpretable results under log-transformation - is simpler, we discuss it first in Subsection 2.1, and then turn to the second - lower and upper quartiles reflecting clusters - in Subsection 2.2. In Subsection 2.3, the asymptotic distributions of our test statistics are presented.

2.1 Level vs. log.

Let L_{Nt}, M_{Nt} and U_{Nt} denote the sample lower quartile, median, and upper quartile of y_{it} respectively, $i = 1, ..., N$, and $t = 0, 1$; define the population versions, respectively, as L_t, M_t and U_t. The subscript i will be often omitted in the remainder of this paper.

The hypothesis of convergence is

$$U_1 - L_1 - (U_0 - L_0) < 0. \tag{1}$$

Suppose now we use $\ln(y_t)$ instead of y_t. Then the population lower and upper quartiles of $\ln(y_t)$ are, respectively $\ln(U_t)$ and $\ln(L_t)$; in general, for a non-decreasing transformation T(y), the α-th quantile of T(y) is T(the α-th quantile of y). The same form of hypothesis, but in log, is

$$\begin{aligned} & \ln(U_1) - \ln(L_1) - (\ln(U_0) - \ln(L_0)) < 0 \qquad (2) \\ \Leftrightarrow \quad & (U_1/L_1)/(U_0/L_0) < 1 \Leftrightarrow (U_1 - L_1)/L_1 < (U_0 - L_0)/L_0. \end{aligned}$$

With log transformation, we are defining convergence as a *declining IQR relative to the lower quartile.* In view of this, it is fitting to call (1) "absolute convergence" and (2) "relative convergence". Hence, within Q-convergence, we get $2 \times 2 = 4$ convergence concepts: conditional or marginal depending on whether x is used or not, and absolute or relative depending on whether y or $\ln(y)$ is used.

What we just showed is that, although the IQR-based inference is not invariant to log-transformation, IQR difference between the two periods renders clear interpretations whether y_t or $\ln(y_t)$ is used. This is in sharp contrast to using mode. To see this, suppose y has a differentiable bimodal density $f(y)$ with the two modes m_1 and m_2. Then

$$z \equiv \ln(y) \text{ has density } g(z) = f(e^z)e^z.$$

Differentiating $g(z)$, the first-order condition is

$$f'(e^z) + f(e^z) = 0.$$

Since $f'(m_1) = f'(m_2) = 0$, we get $f'(e^{\ln(m_1)}) = f'(e^{\ln(m_2)}) = 0$: the first order condition is positive at $\ln(m_1)$ and $\ln(m_2)$. That is, $\ln(m_1)$ and $\ln(m_2)$ are not the modes for the $\ln(y)$ distribution differently from the quantile case; they are now located on upward sloping sections of the $\ln(y)$ density. Furthermore, $g(z)$ is not necessarily bimodal. It is far from clear how to reconcile and interpret inferential results on modes obtained under y and $\ln(y)$.

2.2 Quartiles as cluster representatives.

There are many ways to find clusters in a given sample (see Kaufman and Rousseeuw (1990) e.g.). We will show that using "k-median", the lower and upper quartiles represent the two clusters under one condition. Before this, however, we introduce a couple of facts to motivate the k-median.

Consider a location model $y_i = \beta + u_i$ where u_i is a continuously distributed error term, and β is the parameter of interest. Define the indicator function $1[A] = 1$ if A holds and 0 otherwise. Observe the following: with b ranging over a compact parameter set B,

$$E(y) = \beta \Rightarrow \arg\max_b \ (1/N)\sum_i (y_i - b)^2 \text{ is consistent for } \beta;$$

$$Median(y) = \beta \Rightarrow \arg\max_b \ (1/N)\sum_i |y_i - b| \text{ is consistent for } \beta;$$

$$Mode(y) = \beta \Rightarrow \arg\max_b \ (1/N)\sum_i 1[|y_i - b| < \delta] \text{ is consistent for } \beta \text{ as } \delta \rightarrow 0;$$

see Koenker and Bassett (1978) for the median, and Lee (1989) for the mode.

Now suppose we want to estimate two, not one, location measures in the y-distribution, perhaps thinking that there are two clusters in the data. One well-known way to do that is using one of the following that generalizes the above three cases: with b_1 and b_2 ranging over B,

$$\arg\max_{b_1, b_2} \ (1/N)\sum_i \min\{(y_i - b_1)^2, (y_i - b_2)^2\};$$

$$\arg\max_{b_1, b_2} \ (1/N)\sum_i \min\{|y_i - b_1|, |y_i - b_2|\};$$

$$\arg\max_{b_1, b_2} \ (1/N)\sum_i \min\{1[|y_i - b_1| < \delta], \ 1[|y_i - b_2| < \delta]\}.$$

The minimizers can be called 2-means, 2-medians and 2-modes, respectively, and they reflect the central tendency of the two clusters. Generalizing this to k clusters leads to the names k-means, k-medians and k-modes; see Pollard (1981,1982) and the references therein for the mean and median case.

Differently from the one cluster case, however, the parameters to which the estimators converge in the two cluster cases are not necessarily easily characterized. For the 2-means, the parameters β_1 and β_2 are defined by the first-order conditions of the minimization:

$$\beta_1 = E(y|y < (\beta_1 + \beta_2)/2), \quad \beta_2 = E(y|y > (\beta_1 + \beta_2)/2);$$

here, β_1 and β_2 are hard to interpret. For the 2-median, however, the first-order conditions become

$$P(y < \beta_1) = P(\beta_1 < y < (\beta_1 + \beta_2)/2), \ \ P((\beta_1 + \beta_2)/2 < y < \beta_2) = P(\beta_2 < y),$$

which are satisfied if β_1 and β_2 are the lower and upper quartile, respectively, and *if the median is* $(\beta_1 + \beta_2)/2$. For the 2-modes, if y is bimodal, then the two modes satisfy the first order conditions; otherwise, the estimation becomes degenerate for only one mode can be estimated.

The exact computational details and requisite second order minimization conditions for the k-means and its variations are not needed for us; what the above discussion shows is that the *two quartiles represent the central tendencies of the two clusters if* $M_t = (L_t + Q_t)/2$ *as much as the two modes possibly do*. The condition $M_t = (L_t + Q_t)/2$ is testable with the given data; even if the condition is not met, still Q-convergence retains the two advantages: IQR represents dispersion as SD does, and IQR is more easily interpretable for the level data as well as for the log-transformed data.

The mode-based convergence tests as done in Bianchi (1997) need nonparametric estimators converging very slowly at the rate $N^{1/5}$, which is problematic given the limited number of countries in the world; two other shortcomings of the test are spurious modes at the tails of the estimated y density and sensitivity of the nonparametric estimator to the "bandwidth". These problems may be attributed to the non-smoothness inherent to the minimand for the 2-modes. In contrast, quartiles can be estimated $N^{1/2}$-consistently without any bandwidth, although their asymptotic distributions involve the y density which in turn requires a bandwidth for estimating the asymptotic distribution.

2.3 Asymptotic distributions for test statistics.

As for the asymptotic inference in marginal convergence, observe the following asymptotic expansion:

$$\begin{aligned}\sqrt{N}(L_{Nt}-L_t) &= -(1/\sqrt{N})\sum_i f_t(L_t)^{-1}\{1[y_{it}\le L_t]-0.25\}+o_p(1);\\ \sqrt{N}(M_{Nt}-M_t) &= -(1/\sqrt{N})\sum_i f_t(M_t)^{-1}\{1[y_{it}\le M_t]-0.50\}+o_p(1);\\ \sqrt{N}(U_{Nt}-U_t) &= -(1/\sqrt{N})\sum_i f_t(U_t)^{-1}\{1[y_{it}\le U_t]-0.75\}+o_p(1)\end{aligned} \quad (3)$$

where f_t denotes density of y_t.

Subtracting the first from the last, with $R_{Nt}\equiv U_{Nt}-L_{Nt}$ and $R_t\equiv U_t-L_t$ ("R" from interquartile-Range), we get

$$\sqrt{N}(R_{Nt}-R_t) = -(1/\sqrt{N})\sum_i [f_t(U_t)^{-1}\{1[y_{it}\le U_t]-0.75\}-f_t(L_t)^{-1}\{1[y_{it}\le L_t]-0.25\}]+o_p(1).$$

This implies

$$\sqrt{N}(R_{Nt}-R_t)\Rightarrow N(0,C_t)$$

where $C_t\equiv E[f_t(U_t)^{-1}\{1[y_{it}\le U_t]-0.75\} - f_t(L_t)^{-1}\{1[y_{it}\le L_t]-0.25\}]^2$.

Hence, we can use the following for the Q-convergence hypothesis (1) or (2):

$$\sqrt{N}\{R_{N1}-R_{N0}-(R_1-R_0)\}\Rightarrow N(0,C_{10})$$

where $C_{10}\equiv E[f_1(U_1)^{-1}\{1[y_{i1}\le U_1]-0.75\}-f_1(L_1)^{-1}\{1[y_{i1}\le L_1]-0.25\} - \{f_0(U_0)^{-1}\{1[y_{i0}\le U_0]-0.75\}-f_0(L_0)^{-1}\{1[y_{i0}\le L_0]-0.25\}\}]^2$.

The asymptotic variance of the test statistic C_{10} can be estimated consistently by replacing U_t and L_t with their sample versions, and $f_t(z)$ with a kernel nonparametric estimator, say $f_{Nt}(z)$:

$$f_{Nt}(z) \equiv (1/(Nh)) \sum_i K((y_{it} - z)/h)$$

where $h \to 0$ as $N \to \infty$, and $K(\cdot)$ is a kernel function (e.g., the N(0,1) density); h is the "bandwidth". The bandwidth can be chosen in many ways, but a reasonable thing to do in one-dimensional case is to draw $f_{Nt}(z)$ over z and choose h such that $f_{Nt}(z)$ is not too jagged (if h is too small) nor too smooth (if h is too big). To be specific, a consistent estimator for C_{10} is

$$\begin{aligned}(1/N)\sum_i [\, & f_{N1}(U_{N1})^{-1}\{1[y_{i1} \le U_{N1}] - 0.75\} - f_{N1}(L_{N1})^{-1}\{1[y_{i1} \le L_{N1}] - 0.25\} \\ & -\{f_{N0}(U_{N0})^{-1}\{1[y_{i0} \le U_{N0}] - 0.75\} - f_{N0}(L_{N0})^{-1}\{1[y_{i0} \le L_{N0}] - 0.25\}\}\,]^2\end{aligned}$$

For the quartiles to represent two clusters, we also need to test for

$$G_t \equiv M_t - (L_t + Q_t)/2 = 0. \tag{4}$$

Defining $G_{Nt} \equiv M_{Nt} - (L_{Nt} + Q_{Nt})/2$, from (3),

$$\begin{aligned}\sqrt{N}\{G_{Nt} - G_t\} &= -(1/\sqrt{N})\sum_i [\, f_t(M_t)^{-1}\{1[y_{it} \le M_t] - 0.50\} \\ & -f_t(L_t)^{-1}\{1[y_{it} \le L_t] - 0.25\}/2 - f_t(U_t)^{-1}\{1[y_{it} \le U_t] - 0.75\}/2\,] + o_p(1).\end{aligned}$$

Thus,

$$\sqrt{N}\{G_{Nt} - G_t\} \Rightarrow N(0, C_g) \tag{5}$$

where $C_g \equiv E[\, f_t(M_t)^{-1}\{1[y_{it} \le M_t] - 0.50\} - f_t(L_t)^{-1}\{1[y_{it} \le L_t] - 0.25\}/2 - f_t(U_t)^{-1}\{1[y_{it} \le U_t] - 0.75\}/2\,]^2$.

A consistent estimate for C_g can be obtained, doing analogously to C_{10}.

3 Empirical Analysis.

In the preceding section, we introduced Q-convergence and provided our test statistics. This section applies Q-convergence as well as other convergence concepts to a panel data with $N = 125$ and $T = 5$ (1970, 1975, 1980, 1985, and 1989) drawn from the Penn World Tables 5.6 of Summers and Heston. The response variable is the real GDP per worker (GDP from now on) and $\ln(GDP)$ (LGDP from now on). The selection of the five periods, which are five year apart except the last year, is due to missing values for some variables; including years earlier than 1970 or later than 1989 decreases N nontrivially. In Subsection 3.1, the existing approaches in the literature (β-convergence, σ-convergence, and modal convergence) are applied to the data. In Subsection 3.2, Q-convergence is applied to the data; both the marginal and conditional versions will be examined. In Subsection 3.3, the findings in Subsection 3.1 and 3.2 are put together to ease comparison.

3.1 $\beta-$, $\sigma-$, and modal convergence.

As shown in Barro and Sala-i-Martin (1992, 1995), $\beta-$convergence occurs when economies that start out poor tend to display high growth rates, which means a negative correlation between the growth rate of GDP and the GDP at the base period. More specifically, for the simple least squares estimation (LSE) of GDP growth rate on GDP at the base period, a negative slope coefficient is taken as the evidence for $\beta-$convergence. From σ-convergence viewpoint, however, this way of inference is subject to the "Dalton's fallacy", which is in essence that, although a positive slope is sufficient for $\sigma-$divergence, a negative slope is not sufficient for $\sigma-$convergence. This criticism notwithstanding, $\beta-$convergence is widely used in practice. Doing the simple LSE for our data with GDP (LGDP) as the response variable, the t-value for the slope estimate is -0.53 (-0.19): there is an evidence for β-convergence, but it is not statistically significant.

As for $\sigma-$convergence, as already mentioned, there are various dispersion measures for a given distribution; here, we use the most popular one in the σ-convergence literature: SD. Two upper lines in Figure 1 show the SD of GDP and LGDP divided by the base year SD. While SD(GDP) shows a clear $\sigma-$divergence, SD(LGDP) shows hardly any change although it indicates σ-divergence. To test this formally, let V_{Nt} denote the sample variance for year t: $V_{Nt} \equiv (1/N)\sum_{i=1}^{N}(y_{it} - \bar{y}_t)^2$ where $\bar{y}_t \equiv (1/N)\sum_{i=1}^{N} y_{it}$. Then, under $\sigma-$convergence,

$$\sqrt{N}(V_{N1} - V_{N0}) = (1/\sqrt{N})\sum_{i=1}^{N}\{(y_{i1} - \bar{y}_1)^2 - (y_{i0} - \bar{y}_0)^2\}$$
$$\Rightarrow N(0, C_v) \quad where \quad C_v \equiv E\{(y_{i1} - E(y_{i1}))^2 - (y_{i0} - E(y_{i0}))^2\}^2.$$

C_v can be estimated consistently by

$$C_v \equiv (1/N)\sum_{i}\{(y_{i1} - \bar{y}_1)^2 - (y_{i0} - \bar{y}_0)^2\}^2$$

Applying this test, σ-divergence for GDP is statistically significant, while that for LGDP is not. The t-values for GDP are 4.10, 4.58, 5.22, and 5.57 for 70-75, 70-80, 70-85, and 70-89, respectively. For LGDP, they are 0.65, 0.85, 1.03, and 2.23, respectively.

Along with GDP and LGDP, the GDP normalized by the year's GDP sum has been also used in the literature, and this yields significant $\sigma-$convergence as shown in the third line of Figure 1, which is opposite to that of SD(GDP); Dalgaard and Vastrup (2001) explain analytically why this can happen. Within the $\sigma-$convergence framework, it is not clear how to reconcile these differences.

Turning to modal convergence, Bianchi (1997) estimated GDP density nonparametrically, and tested for the number of modes in 1970, 1980, and 1989. For GDP, unimodality hypothesis was not rejected at a 5% level of significance, whereas it was rejected for 1980 and in 1989. For LGDP, unimodality was not rejected for all three years. As in $\sigma-$convergence, it is not clear how to interpret this difference across level

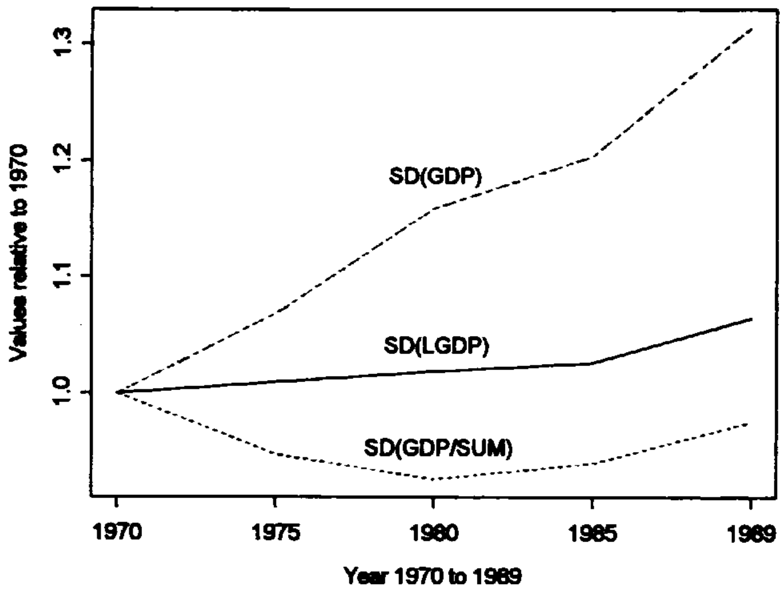

Figure 1: Standatd Deviations relative to 1970

and log data. Also, the test used by Bianchi is rather sensitive to the bandwidth in the kernel estimation; too small a bandwidth can easily yield multiple modes particularly in the tail areas of the density, whereas too big a bandwidth will render an unimodal density even when the density has multiple modes. Bianchi employed a bootstrap to avoid the arbitrariness, but the consistency of the bootstrap procedure is yet to be proven. Since our data is basically the same as Bianchi's data, we show this problem in the following.

In Figure 2, four boxes of kernel density estimates with the N(0,1) kernel are presented for year 1970 and 1989. The left two are for GDP whereas the right two are for LGDP. In the top two boxes, the bandwidth was a "rule of thumb" bandwidth $h_o = 0.9\sigma_y N^{-1/5}$ where $\sigma_y \equiv SD(y)$ for the given year; h_o has been used e.g. by Jones (1997) and shown to be optimal under certain conditions by Silverman (1986). In the bottom two boxes, the bandwidth was a "cross-validation (CV)" bandwidth h_{cv}; there are several ways to do CV, and we used the one minimizing

$$CV(h) = \int \hat{f}(y : h)^2 dy - \frac{2}{N} \sum_{i=1}^{N} \hat{f}(y_i : h)_{-i}$$

where $\hat{f}(y : h)$ is the kernel density estimate evaluated at y, $\hat{f}(y_i : h)_{-1}$ is the "leave-one-out" estimate (i.e., i-th observation is not used for $\hat{f}(y_i : h)$), and $CV(h)$ is minimized over a range for h.

For GDP, the two density estimates for 1970 look unimodal while the two for 1989 look bimodal, but the second mode near the upper tail may be dismissed as an artificial "blip"; for GDP, h_o and h_{cv} are close to each other, and thus both yield similar results. For LGDP, h_o turned out to be much smaller than h_{cv}; h_o renders single mode for 1970 and double modes for 1989, whereas h_{cv} renders only single mode for both years.

3.2 Q–convergence.

Although IQR itself does not depend on any bandwidth, our test statistics do, for their asymptotic distributions include density components. But our tests are much less sensitive to bandwidths than the modal convergence test is; for this, we will show the values of the two bandwidths h_o and h_{cv} along with the values of the test statistics corresponding to them.

The values of the two bandwidths are in Table 1. For GDP, other than in 1970, $h_{cv} < h_o$; for1970 and 1980, $h_o \simeq h_{cv}$, while h_o and h_{cv} are quite different for the other three years. For LGDP, h_{cv} is much bigger than h_o for all years. It can be said that one bandwidth is somewhat "under-smoothing" while the other is "over-smoothing"; using two bandwidths, with one under-smoothing and the other over-smoothing, is a reasonable thing to do when reporting nonparametric estimation results.

Recall that, among the three advantages listed for Q-convergence, the second—two quartiles representing two clusters—requires $G_t = M_t - (L_t + U_t)/2 = 0$. So, before we test for Q-convergence, we need to test for this condition first. Table 2 reports the test statistic value ("t-value") and the two bandwidths h_o and h_{cv}. Despite the

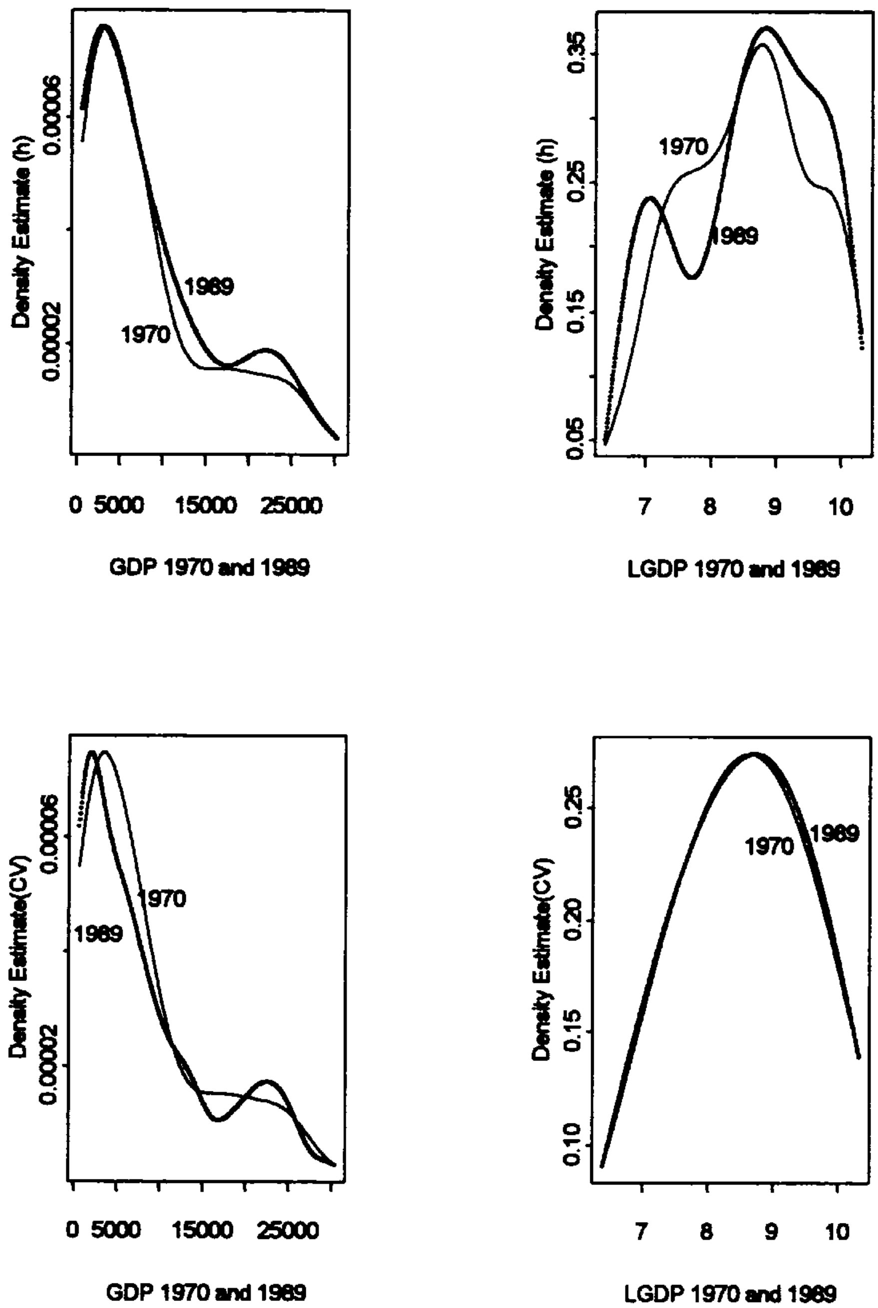

Figure 2: Kernel density estimates in 1970 and 1989 for the GDP and LGDP, using two bandwidths, rule-of-thumb (top panels) and CV (bottom panels), respectively.

Table 1: Bandwidths for GDP and LGDP

	GDP		LGDP	
	h_o	h_{cv}	h_o	h_{cv}
1970	2578.75	2713.70	0.3434	0.9685
1975	2756.04	1669.75	0.3468	0.6593
1980	2987.75	2965.12	0.3501	0.7475
1985	3103.99	1880.55	0.3524	0.9261
1989	3389.48	2157.90	0.3656	1.0568

Table 2: Cluster Representation Tests

	$\|G\|/M$	Test stat.(h_o)	Test stat.(h_{cv})
		GDP	
1970	-0.216	-1.81	-1.84
1975	-0.088	-0.80	-0.68
1980	-0.182	-1.59	-1.58
1985	-0.211	-1.84	-1.74
1989	-0.219	-1.81	-1.83
		LGDP	
1970	0.008	0.76	0.62
1975	0.027	2.51	2.27
1980	0.017	1.52	1.37
1985	0.016	1.42	1.23
1989	0.022	1.86	1.53

Table 3: DIQR and Q-convergence Tests

	DIQR	Test stat.(h_o)	Test stat.(h_{cv})
		GDP	
70-75	2050	1.63	1.62
70-80	3680	2.62	2.62
70-85	3935	2.60	2.68
70-89	5230	3.09	3.23
		LGDP	
70-75	0.1787	0.60	0.56
70-80	0.1518	0.49	0.47
70-85	0.2055	0.65	0.64
70-89	0.3601	1.14	1.08

big differences between h_o and h_{cv}, the test statistic values are quite close, and there is no instance of the conclusion getting reversed due to the bandwidth variation at the size 5%. The null hypothesis that the quartiles represent two clusters cannot be rejected except for LGDP in 1975; even in this case, $|G_t|$ is not big, being only a small fraction (0.027) of M_t.

Turning to Q-convergence tests, Figure 3 presents two boxes with confidence intervals attached to difference of IQR (DIQR) which are the dots in the middle. The upper box is for GDP while the lower box is for LGDP; only h_o is used for Figure 3. The test statistic values for both bandwidths h_o and h_{cv} are provided in Table 3; the test statistics are not sensitive to bandwidth choice. For GDP, other than 70-75, Q-divergence is significant; even for 70-75, the t-value is not small. For LGDP, for all years, Q-divergence is not significant, although the sign of the test statistic indicates Q-divergence. These two different conclusions, however, can be easily interpreted in stark contrast to the other convergence concepts: IQR has increased significantly over the years, but the IQR relative to the lower quartile has increased little. That is, income disparity between two clusters has increased in the absolute term, but not in the relative term, for the lower group's income has increased as well.

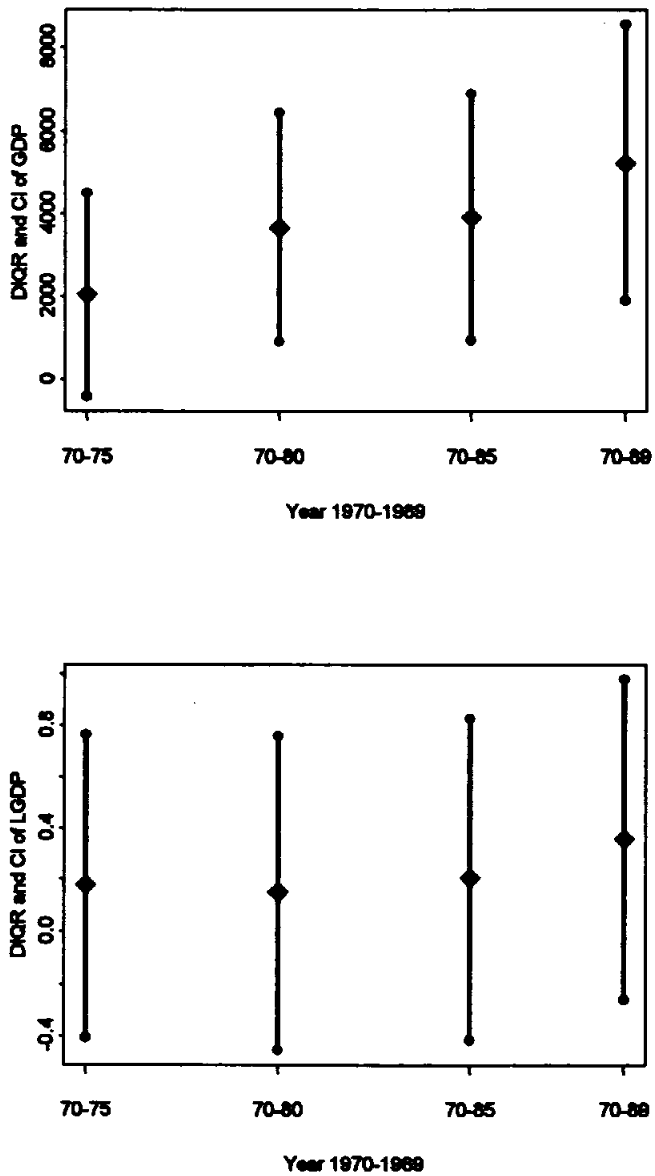

Figure 3: DIQR and 95% Confidence Interval of GDP (top panel) and LGDP (bottom panel) with the rule-of-thumb bandwidth.

Turning now to the conditional Q-convergence, to control for variables relevant for steady state, suppose now

$$\begin{aligned} y_{it} &= x_{it}'\beta_t + u_{it}, \quad \text{the first component of } x_{it} \text{ is } 1, \\ & \quad u_i \text{ is independent of } x_{it}, \quad u_{it} \text{ is iid across } i \text{ at a given } t, \end{aligned}$$

for a parameter vector β_t and error term u_{it}. Estimate β_t by, say, LSE b_{Nt} to get the residuals, $\widehat{u}_{it} \equiv y_{it} - x_{it}'b_{Nt}$. Now proceed as in marginal convergence, replacing y_{it} with $\widehat{u}_{it}$. There are however two differences from the marginal convergence. First, we impose the independence of u_{it} from x_{it} to assure that the location normalization of u_{it} does not matter (for the LSE, the normalization is $E(u_{it}) = 0$); the normalization constant, whatever it may be, is cancelled in IQR due to the differencing. Second, more importantly, since we are using the residuals, not the true error terms, $\widehat{u}_{it} - u_{it}$ is likely to affect the asymptotic distribution. Accounting for this, although not impossible, seems to go beyond the scope of this paper. For the conditional Q-convergence, the asymptotic confidence intervals used should be deemed ad-hoc.

With the conditioning variables in Barro (1991) except literacy rate, Table 4 shows the two bandwidths h_o and h_{cv}, and Tables 5 reports DIQRs with the residuals and the t-values for the two bandwidths. To be specific, the independent variables are enrollment rate in primary, secondary and tertiary schools, student-teacher ratio in secondary and tertiary schools, and dummy variables for OECD and sub-Saharan Africa and Latin America. The variables are from the online data of UNESCO. For GDP, except for 70-85 and 70-89 with h_o, the conditional Q-divergence is not significant. For LGDP, no significant results are obtained for all periods and for both bandwidths. The evidence of Q-divergence for GDP has been weakened by conditioning on the regressors, and the tests have become sensitive to bandwidth choice with this weakening. Due to this problem that also holds for other convergence concepts (controlling for x_{it} almost always dilutes convergence), in the following subsection, we will compare the convergence concepts only for their marginal versions.

Table 4: Bandwidth for Conditional Q-convergence

	GDP.resid		LGDP.resid	
	h_o	h_{cv}	h_o	h_{cv}
1970	1.5577	4.2912	0.1993	1.0072
1975	1.5292	3.4951	0.1812	0.9157
1980	1.7545	1.4015	0.2101	1.0642
1985	2.1832	1.7214	0.1873	0.9271
1989	2.4388	6.7459	0.1813	0.9180

Table 5: DIQR and Conditional Q-convergence Tests

	DIQR.resid	Test stat.(h_o)	Test stat.(h_{cv})
		GDP	
70-75	0.4220	0.67	0.46
70-80	0.5257	0.77	0.64
70-85	1.3598	2.01	1.67
70-89	1.6762	2.15	1.42
		LGDP	
70-75	-0.0231	-0.25	-0.14
70-80	0.0358	0.38	0.22
70-85	0.0369	0.43	0.24
70-89	-0.0754	-0.90	-0.49

Table 6: Various Marginal Concepts

		sign/direction	test
β-convergence:	GDP	convergence	insignificant
	LGDP	convergence	insignificant
σ-convergence	GDP	divergence	significant
	LGDP	divergence	insignificant
modal convergence	GDP	divergence	significant
	LGDP	divergence	insignificant
Q-convergence	GDP	divergence	significant
	LGDP	divergence	insignificant

3.3 Comparison of findings.

Table 6 summarizes the conclusions from the various marginal convergence concepts. Since we did not conduct any modal test for modal convergence, the entries for modal convergence in Table 6 are based upon Bianch (1997)'s tests.

This summary makes a good sense for Q-convergence. *Since IQR represents both dispersion and two clusters of the income distribution, as σ-convergence and modal convergence indicate divergence of GDP, it is natural that Q-convergence does so as well;* here, β-convergence is an "odd man out". Once log is taken on GDP, still divergence seems to hold other than for β-convergence, but this is not significant anywhere.

4 Conclusions.

In this paper, we introduced a new convergence concept "Q-convergence" which is based upon whether interquartile range (IQR) shrinks (convergence) or expands (divergence). Compared with the other convergence definitions, Q-convergence has advantages of reflecting both dispersion and two clusters of the income distribution; also IQR is insensitive to outliers and equivariant to log-transformation, leading to robust inference and easier interpretation of different results using level and log data. A panel data drawn from the Penn World Tables was analyzed and we found significant divergence using the level GDP and insignificant divergence using the log GDP: the income gap between the poor and rich countries has increased, but in relative terms as reflected in the log GDP, the widening gap was rather small and insignificant; this is because the poor countries income has also increased as the gap widened. Convergence or divergence is a matter of definition because one aspect of an economy may converge while another may diverge. The point is then using a definition that is appealing in some senses. We showed that the new concept Q-convergence has this appeal, while reaching conclusions not deviating far from the existing findings in the literature under $\sigma-$ and modal-convergence.

Reference

Barro, R. J., 1991, Economic growth in a cross-section of countries, *Quarterly Journal of Economics*, 106, 407-443.

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

Barro, R. J., and X. Sala-i-Martin, 1995, *Economic growth*, McGraw-Hill.

Bianchi, M., 1997, Testing for convergence: evidence from non-parametric multimodality tests, *Journal of Applied Econometrics,* 12, 393-409.

Dalgaard, C., and J. Vastrup, 2001, On the measurement of $\sigma-$convergence, *Economics Letters, 70*, 283-287.

Durlauf, S.N. and D. Quah, 1999, The new empirics of economic growth, in Handbook of Macroeconomics, volume 1, edited by J.B. Taylor and M. Woodford, 234-308.

Jones, C.I., 1997, On the evolution of the world income distribution, *Journal of Economic Perspectives*, Summer 1997, 11, 19-36.

Kaufman, L. and P. J. Rousseeuw, 1990, *Finding groups in data: an introduction to cluster analysis*, Wiley, New York.

Koenker, R. and G. Bassett, 1978, Regression quantiles, *Econometrica,* 46, 33-50.

Lee, M.J., 1989, Mode regression, *Journal of Econometrics,* 42, 337-349.

Pollard, D., 1981, Strong consistency of k-means clustering, *Annals of Statistics,* 9, 135-140.

Pollard, D., 1982, A central limit theorem for k-means clustering, *Annals of Probability,* 10, 919-926.

Quah, D., 1993, Empirical cross-section dynamics in economic growth, *European Economic Review*, 37, 426-434.

Quah, D., 1996a, Empirica for economic growth and convergence, *European Economic Review,* 40, 1353-1375.

Quah, D., 1996b, Twin peaks: growth and convergence in models of distribution dynamics, *Economic Journal,* 106, 1045-1055.

Silverman, B.W., 1986, *Density estimation for statistics and data analyses*, Chapman and Hall.