

# The World Distribution of Human Capital, Life Expectancy and Income: a Multi-dimensional Approach

Christian Morrisson, Fabrice Murtin\*

September 28, 2005

## Abstract

**JEL classification:** D31, E27, F02, N00, O40.

**Keywords:** Inequality, human capital, multidimensional inequality.

## 1 Introduction

The revival of interest in the world distribution of income has given rise to several articles since the late 1990s. Although there is a consensus about income being the main dimension of economic well being, it is not the only one. Any analysis of world inequality should take into account other components of welfare. For this reason, Bourguignon and Morrisson (2002) introduced national estimates of life expectancy since 1820 in order to estimate the between countries inequality in longevity around the world. Most people consider longevity and education as the two most interesting components of welfare, after income. But until now there was no estimate of inequality in education. The main contribution of this paper is the estimate of inequality in education, within and between countries, in the long run, from 1870 to 2000. Moreover this estimate is combined with estimates of income and life expectancy inequalities in order to provide a multidimensional assessment of world inequality in human development and of its evolution from 1870 to 2000.

Such an enterprise has two advantages. We obtain a comprehensive vision of world inequality and we show that the three evolutions of inequalities in income, education and longevity are completely different. After a large increase until 1910, income inequality has risen very slowly before decreasing between 1980 and 2000. Inequality in education in 1870 was very high, much more than income inequality at any time, but it has decreased steadily until 2000 and is now

---

\*Morrisson: Université Paris I ; address: 36 Chemin Desvallières, 92410 Ville d'Avray, France. Murtin (corresponding author): LSE, London, PSE (joint research unit CNRS-ENS-EHESS-ENPC) and CREST (INSEE), Paris; e-mail: fabrice\_murtin@hotmail.com, address: 12 rue Godefroy, 75013 Paris, France.

much lower than income inequality. Lastly inequality in life expectancy (only the between countries component can be measured) has been always moderate. Inequality in life expectancy has increased from 1820 to 1930 and fallen considerably from 1930 to the present day.

In section I this paper looks at the methodology and the data. Section II presents the overall evolution of world education inequality since 1870. Section III focuses on education inequality between countries and regions. Section IV provides a synthesis of income, longevity and education inequalities with a multidimensional approach.

## 2 Methodology and data

We applied the same methodology than Bourguignon and Morrisson (2002) : here human capital, measured by the number of education years, replaces GDP per capita<sup>1</sup>. Before estimating human capital, we updated the figures on GDP per capita and population, adding 2000 and using the last estimates of Maddison (Maddison 2003). For education, estimates of average human capital (the number of years of schooling) were assembled in 91 countries from 1870 to 2000. More precisely we associated two bases, the first for 1870-1960 being a new one whereas the second (1960-2000) is given by Cohen and Soto (2003). We have followed the same methodology that Cohen and Soto in order to obtain a consistent set of data for a 130 years period (cf. Annex 1). In each country, we gave preference to census results over the series of Mitchell on school enrollments. When differences arise between the two sources, we rectify these series. It is impossible to estimate the human capital before 1870 because we need enrolment data since 50 years in order to obtain the school attainment of the population aged of 15 to 65 years.

The other source could be a census in 1830 or 1850 which gives such information, but there is no census before 1870. In Western European countries statistics of school enrolment are available since 1820-40, but in other countries we cannot find any statistics before the end of the 19th century. For the less developed countries, the series concerning enrolment begin often in 1920 or later. In these countries we have assumed a steady growth of the enrolment rate starting from a minimum which is very low in 1820. Simulations with thrice the latter value provide very close estimates of national human capital. Our series concern only primary and secondary education. If we exclude the U.S. and the Canada, the enrolment rate in universities were very low until 1950. On the other hand the estimates of Cohen and Soto include university education which is taken into account in our basis since 1960.

As in Cohen-Soto (2001) we compute the average number of primary and secondary years of schooling by inferring the enrolment rates for each cohort of

---

<sup>1</sup>The number of education years is only a proxy for human capital, because we measure the quantity but not the quality of education. As this quality can vary a lot from a country to another, the amount of human capital obtained after the same number of years can differ (cf. World Bank 2005 Chapter 2).

age at each date. This is made possible because the series from Mitchell provide the number of children in primary and secondary schools as well as population by age. Usually age pyramids were available each ten years and missing data has been interpolated. To achieve the computation we also needed some information on schooling's duration, dropout and repetition rates. While we used Unesco data for the two latter variables, completed primary and secondary were both assumed to last six years<sup>2</sup>. The average human capital in primary and secondary was then deduced from these duration assumptions and from enrolment rates of each cohort of age between 15 and 65 in the population. Annex 1 mentions the sources that have been used, and Annex 2 describes in details the procedure to infer mean human capital of primary and secondary schooling.

Then we used another dataset on illiteracy rates in order to improve our description of the distribution of years of schooling. This second database enabled us to refine the assumptions on the mean duration of schooling and to correct for measurement errors in Mitchell's data. Indeed, given the calculated stocks  $H^P$  and  $H^S$  of primary and secondary schooling and the percentage  $p^I$  of illiterates in the population, we can infer the percentage  $p^P$  of the population displaying only primary schooling, the percentage  $p^S$  of the population displaying primary and secondary schooling, and the duration  $h$  of primary and secondary cycles that are assumed to be equal. These are given by

$$\begin{cases} H^P = (p^P + p^S) h \\ H^S = p^S h \\ p^P + p^S + p^I = 1 \end{cases}$$

This two-steps method, first computing aggregate stocks of human capital, second adjusting its distribution in three groups of education, is likely to draw a realistic picture of both the mean educational attainment in the population and the duration of attained schooling, in absence of any information on countries' educational system.

Inequality indices are computed on the distribution of these  $3 \times 33 = 99$  groups. All the groups are pooled and ranked according to the number of years of education and then the cumulative function and Lorenz curve of the world distribution of education is computed. As illiteracy is the common rule for most countries before 1960, we used a corrected Gini index as in Thomas-Wang and Fan (2001).

Besides, we have smoothed the cumulative distribution function in each country to transform a three groups distribution into a continuous one. Any smoothing relying on kernel techniques is likely to seriously bias the mean value of human capital for some countries because kernels drive poor estimates of the density function with only three observations. On the contrary, the smoothing

---

<sup>2</sup>This hypothesis is a rough estimate that we can use because there is no detailed information on the lengths of primary and secondary schooling in each country from 1870 to 2000. The length varies according to the country and the period. For example in present day France the respective lengths are 5 (primary) and 7 (secondary) ; but until 1950, the two lengths were equal (the pupils engaged in secondary schooling left primary school after 5 years, but the others who represented a large majority remained in primary school 7 years).

procedure we applied to the cumulative distribution function has not changed the mean human capital of any country at any date from more than one percent of its original value<sup>3</sup>. Computed inequality indices on the smoothed distribution were close to those corresponding to the three-groups distribution until 1970, then they were somewhat higher because of a non-negligible dispersion within primary and secondary educational groups.

In conclusion, using the data collected in 90 countries, we estimated the average human capital in each country or group of countries of the AER 2002 basis. Each country or country group represents at least 1 per cent of world population or world GDP in 1950. All countries which are important are considered individually. To allow a simpler analysis, these countries or country groups were aggregated in 6 blocks, defined geographically, historically or economically : Africa, Asia, excluding Japan, Korea and Taiwan, these 3 countries, Latin America, excluding Argentine and Chile, Eastern Europe (which include all the countries of the ex-USSR) and Western Europe (including Austria, Hungary and Czechoslovakia) and its offshoots in America (Canada, U.S., Argentine, Chile) and in the Pacific.

### **3 Evolution of world distribution of education since 1870**

Tables 1 a and 1 b depict the world inequality in income, life expectancy and education. Table 1 a corresponds to the three groups distribution of education, while Table 1 b relies on the continuous education distribution. The former constitutes a lower bound for inequality, while the latter is an upper bound. As education is not likely to be scattered continuously, the real inequality is probably close to estimates of Table 1 a. It shows an exceptional inequality in 1870 with a Gini coefficient reaching 0.79 and a Theil index of 1.4. Such inequality is far higher than income inequality in any time. The world in 1870 was characterized by a huge gap between the literate and illiterate populations which we hardly can think of today.

If we exclude the elite in Eastern Europe and China or Japan where around 40-50% of men were literate, education was the monopoly of Western European populations, who lived in Western Europe, in its offshoots, and as a minority in less developed countries, mainly in Latin America. There is a main difference between income and education : a minimum income is necessary in order to survive, but many people can have no any human capital if there are illiterate and survive nevertheless. In 1870 around  $\frac{3}{4}$  of the world population is illiterate and has a human capital equal to 0. This fact explains the gap between the share of 80% of the world population in the world stock of human capital : 11% and the share of the top quintile, nearly 90% (instead of 60% of the world income).

---

<sup>3</sup>we depict in Annex 2 the smoothing procedure.

We cannot estimate precisely the world education distribution at the beginning of the 18th century. But it is sure that inequality was much lower than in 1870 on account of two facts. First, in the intermediary deciles, the literacy rate in men population of China and Japan was nearly the same than in 1870. But in the top quintile the situation was completely different : the literacy rates in Western Europe, except in England, Scotland, Netherland and Sweden, were very low, always less than 20%. For example in 1700 this rate was probably lower in France than in China. The expansion of primary education to nearly the whole population during the 18th and the 19th centuries in Western Europe and its offshoots, explains a large increase of inequality and an inverted U-shape for world education distribution as for world income distribution (if we assume that the decrease of inequality since 1980 will go on).

But the evolutions of the two curves are very different : education inequality has decreased steadily since 1870 whereas income inequality has reached its maximum in the middle of the 20th century and lessened only slightly since 1980. As a result income inequality today is higher than in 1870, but education inequality is much lower. Between 1870 and 2000, the Gini coefficient decreased by 50%, the Theil index and the mean logarithmic deviation by 77% and 60%. The consequence is an important difference in 2000 : the Gini coefficient and the Theil index for education are 40% and 60% lower than for income. The world education inequality is moderate whereas the world income inequality remains very high. The distribution of human capital shows a limited gap between the bottom 60% and the top decile. The share of the bottom 60% reaches 32% (instead of 14% of income), and the share of the top decile is 22%. So the average human capital of this decile amounts to four times the average one of the 60% poorest (instead of 22 times more for income).

This evolution results partially of the characteristics of human capital. Even if the expansion of post-secondary education in developed countries increases the average years of schooling , there is a limit. In a country where everyone has access to such education, the maximum for this average would be 18, which represents no more than three times the average in a country where the enrolment rate in primary education is 100% and 0% in secondary and post-secondary education. On the contrary there is not such a limit for income differences between countries.

As the average human capital of the top decile in 2000 is around 15 years, it is near to the limit of 18 years. This figure, 15, represents only the double of the world average (6.7 years more precisely). We observe that the average human capital and the average income have increased similarly since 1870 : they are multiplied by 6.3 and 6.7<sup>4</sup>. There is also a concordance between illiteracy rate and extreme poverty (less than 1 dollar a day). The illiteracy rate has decreased since 1870 from 75% to 22% and the extreme poverty from 75% to 20%. So the evolutions of these two essential indicators (world average and percentages of people who have no access to education or a minimum income)

---

<sup>4</sup>We can compare our estimate to another one only for 1960 -2000. The variation of the average human capital in World Bank (2005) and in Table 1a are quite similar: +86% and +72%.

are parallel and they show an improvement which never happened before in mankind history. But it is only for education that the increase of average human capital is combined with a strong decrease in inequality. The distribution is much less unequal today than in 1870 and in the same time the average capital has been multiplied by 6.3.

The decomposition of human capital inequality in two components : the inequality within countries (or country groups) and the inequality between countries (or country groups) shows another difference with income inequality. The within component of inequality has decreased much more than the within component of income inequality (less 82% instead of less 12% for the Theil index). Moreover the between component has fallen rapidly : the Theil index in 2000 is only 0.087 instead of 0.590 in 1870. For 1960-2000 an estimate is given by World Bank (2005) which does not take into account the weighting by population. Despite this difference, we observe similar decrease of the Theil Index: -60% (World Bank 2005), -70% (Table1a). In total inequality, the contribution of the between component plays only a marginal part: 27% in 2000 for the Theil index and 7% for the MLD, two figures in agreement with The World Bank (2005) estimate (less than 20%). It is exactly the contrary for the income inequality between countries : the increase of the Theil index has reached +163%. The gap between the poorest region (Africa) and Western Europe for the average human capital is only 1 to 4 instead of 1 to 12 for the average income.

This fall of inequality between countries is the result of the extension of primary schooling in a large majority of countries (except in Pakistan, the north of India and several sub-saharian African countries where enrolment rates of girls are often much lower than those of boys). This expansion entails two effects. The within inequality decreases because the average human capital of the bottom deciles rises from 0 to 4 or 5 years when the enrolment rate in primary school exceeds 80% since 30-40 years. The between inequality decreases also because the national average reaches 5 years or more instead of 1 or 2 years when the enrolment rates in primary school is below 30 /40%. Finally this expansion reduces drastically the illiteracy rates. When countries give priority to primary schooling , as the Western European ones did during the 19th century, the decrease of within inequality is much more rapid than when the governments extend secondary and post secondary-schooling and leave 30 or 50 % of children with no access to primary schooling as is the case in several African countries (cf. Berthélemy 2004).

## 4 A regional approach to education inequality

The decomposition of world population in 6 blocks allows a better understanding of the changes in the world distribution of education since 1870. Table 3 shows the composition of 3 quantiles : the bottom 80%, the 9th and 10th deciles between the 6 blocks (the first line gives the population distribution). The main factors which explain the variations from one year to the other are the different

rates of growth of average human capital and of population (the shares in world population of Latin America and Africa respectively, have been multiplied by 3 and 2 between 1870 and 2000, whereas the shares of Western Europe and Eastern Europe have decreased).

The two opposite blocks are Western Europe and Africa. In 1870, Western Europe (and offshoots) had an edge on the rest of the world, which increased until 1910. At this time the share of Western Europe in the top decile (which had an average human capital equal to 8.4 years of schooling) reached 74%. It was more than the share of the same region in the top income decile (64%). If we consider that secondary schooling is the condition of access to technology, in 1910, Western Europe had in some respect the quasi-monopoly of advance in knowledge and technology. Today this monopoly has disappeared. The share of Western Europe in the top decile (nearly 15 years of schooling which implies post-secondary education) is only 33%, the same that the share of Asia (excluding Japan, Korea and Taiwan). If we include these 3 countries, the share of Asia prevails with 42%. Extrapolating these trends we can foresee that in a few years time Asia (defined geographically) will attain 50% and Western Europe less than 30%, which will entail important consequences in the world distribution of scientific and technological supremacy. Moreover there is a significant contrast between the share of Western Europe in income top decile and education top decile. The income share is around 65% , the same than in 1910, whereas the education share is 33% after a large decrease since 1910. So Western Europe has kept an important advantage in world income distribution in spite of losing its supremacy in education.

The African case is a counter-example. First it is today the poorest region in the world. The ratio education/population for the top decile share is only 0.4 (instead of 0.62 in Asia) ; but this handicap is not new. In 1870, the share of Africa in the 9th and 10th deciles were 0.6 and 0.1, and, taking into account the differences of populations, the share of Asia in the top decile was 30 times higher than the African share. This was a legacy of the past : at the beginning of the 19th century nearly all African populations were illiterate (except the Arab population in the north of Africa) whereas nearly 50% of Chinese and Japanese men could read and write. In 2000, the shares of Africa and Western Europe in world population are the same. But the percentage of Western Europe in the top education decile is 6 times higher. Nevertheless the African situation was worse before. In 1980 Western Europe share (for the same population) was 12 times higher. Even if the situation remains unfavourable, Africa is slowly catching up with the rest of the world. We must remember the situation in the 19th century in order to understand better its present lagging behind.

Of course the success story of world education is the Japan, Korea and Taiwan case. In 1870 the shares of these countries in the 9th and 10th deciles were more or less alike the shares of Eastern Europe. In 2000, these shares are the same that the Western Europe shares, if we take into account the population effect. It is the only group of countries which has caught up completely Western Europe. The situation of Latin America and Eastern Europe, at same population share, has improved, particularly in Eastern Europe, but the gap with

Western Europe has not disappeared.

Despite the progress of Asia and a large decrease of world education inequality since 1910, the geographical concentration of people who have less education than the world average has increased. The shares of Africa and Asia in the bottom 80% of world distribution has increased from 66% in 1910 to nearly 81% in 2000. This is slightly more than the relative population increase (68% instead of 57%). We can explain such a result by the primary education expansion in other regions where illiteracy has disappeared.

The table 4 gives a geographical view of education inequality, with the average human capital by region, total and primary (the difference equals the human capital in secondary) and the illiteracy rates. We have splitted Asia in three items: China, South Asia (India plus Bangladesh, Burma and Pakistan), and the rest of Asia. This decomposition is necessary because Asian population amounts to more than 50% of world population in 2000 as in 1870.

We observe in 1870 three education levels : the lowest (less than 0.2 year) in Africa and Asia (except China), an intermediary one : Latin America, Eastern Europe, Japan, Korea and Taiwan (and offshoots) which exceeds 3 years. The illiteracy rate exceeds 90% in the first group, 70% in the second one, whereas 70% of West Europeans are literate. So there is a huge gap between the first group and Western Europe : more than 1 to 20 (3.16 years instead of 0.15). An important point is the advance of China. The estimate, one year, is a rough approximation. It could be 0.9. But this last figure does not change the lag of other Asian countries (including Indian empire) and Africa. On the other hand the human capital in Japan was higher than in Korea and Taiwan. So, in 1870, there were only two countries in the world which reached 0.8-1 year outside Europe (in Latin America education was the monopoly of European population). In fact this figure means that around 40% of men and 5% of women could read and write 1500 graphic signs, which demands about 3 or 4 years of schooling. A small minority only knew several thousand signs after 6 or 8 years of schooling. As the average capital in China and Japan was the same at the beginning of the XVIII<sup>e</sup> century, these countries were the only one in the world who had the same average human capital than Western Europe three centuries ago.

Since 1870 the huge gap between the first and the third group has disappeared. In 2000, we find in the first group Africa and India only because the average human capital in other Asian countries has increased much more than in India. The human capital is 4 years in this group instead of 12 years in Western Europe and the ratio has been reduced from 1/20 to 1/3. Moreover Japan, Korea and Taiwan have caught up Western Europe. In the second group the average human capital varies between 6 years in China and 9.7 years in Eastern Europe. So Western Europe has lost its advance. Japan, Korea and Taiwan have the same capital and the difference with Eastern Europe is small. Finally the differences between Western Europe and some other Asian countries (which have human capital higher than 6.7), China, Latin America are no longer important. These changes explain an important reduction of world inequality in human capital and the end of the education comparative advantage of Western



Europe in world competition. Illiteracy which dominated in 1870 with rates exceeding 70% everywhere except in Western Europe is now a regional problem. It remains important only in Africa, more precisely in Sub-Saharan Africa and India (including Bangladesh, Burma and Pakistan) with rates around 50%.

For illiteracy reduction as for human capital accumulation, primary schooling is the key factor. In 2000, the human capital in secondary schooling exceeds two years only in Latin America, Eastern Europe, Western Europe and Japan, Korea, Taiwan. This fact is unexpected because we observe today the expansion of secondary schooling everywhere. But the average secondary schooling in 2000 depends of enrolment rates since 1950 and we have forgotten the low rates in the 50', the 60'. In 2000 secondary human capital is important (more than 3 years) only in Japan, Korea, Taiwan and Western Europe where the expansion of secondary schooling began in the 50' or before. Until 1980, secondary human capital was lower than 1.7 year everywhere except in these two blocks. From the early XIX<sup>o</sup> century to 1980, the differences in primary schooling have induced the main regional inequalities in human capital. But in the next years, the expansion of secondary schooling will become the key factor.

## 5 Inequality in human development since 1870

In this section we aggregate the three available dimensions (GDP per capita, education, life expectancy) to construct a one-dimensional index of social and economic development. We propose three different methods of aggregation, and then we compute inequality in this resulting human development index.

From a practical perspective, assessment of multi-dimensional inequality can be achieved with two strategies: one is to look at Lorenz-dominance criteria on the joint distribution of income, education and life expectancy<sup>5</sup>; the other is to derive an index that will depict the joint evolution of the latter attributes. As it is plausible that multi-dimensional inequality has not evolved monotonically because inequalities in education and income have evolved in opposite direction across the 20th century, we focus on the second approach<sup>6</sup>.

To aggregate the various attributes into a single index, one difficulty to overcome is the choice of the aggregation function and that of the relative weight

---

<sup>5</sup>Following Atkinson-Bourguignon (1982), it is clear that first-order Lorenz-dominance criteria will not be fulfilled for all utility functions: for instance, the marginal distributions of income observed in the 19th century Lorenz-dominate those observed in the 20th century, while it is the opposite for education. This is sufficient to exclude all utility functions with negative cross-derivatives, which are the most plausible assumptions, from the set of candidates for which first-order dominance could be assessed.

<sup>6</sup>Becker et al. (2005) have computed a "full" growth rate that includes longevity gains for the period 1960-2000. The value of longevity gains in terms of annual income is added to the GDP per capita. As the life expectancy gains have been much more important in other regions (if we exclude Sub-Saharan Africa) than in Europe and North America, the decrease of cross-country inequality between 1960 and 2000 is important for "full" income instead of small for GDP per capita. If welfare is affected by quantity of life represented by longevity, on the contrary assuming that the number of years of education increases welfare, excluding the incidence on GDP per capita, is an open question.

of the different attributes. A popular procedure has been suggested by Ram (1982) and Maasoumi (1989), where natural candidates are derived: it consists in using a weighted arithmetic mean of attributes where the weights are the elements of the first eigen vector in the principal component analysis of the data. The idea at the core of this procedure is that the first principal component catches the largest part of the variance in the original variables. Thus it provides a linear combination of attributes that encompasses most of the information that is contained inside. Indeed, in our applications the first principal component catches between 70% and 90% of the data's variance, depending on the retained index.

Nevertheless there still remains another difficulty inherent to the exercise of aggregating variables of different scale. As the scaling of each attribute has theoretically an influence on the computed weight of the principal component analysis, we retain three different methods of rescaling. If  $Y_{i,t,k}$  is the value of attribute  $k$  in country  $i$  at date  $t$  and  $\alpha_k$  the weight of attribute  $k$  given by the principal component analysis, we define the following three indicators

$$\begin{aligned} I_1 &= \sum_k \alpha_k \frac{Y_{i,t,k} - Y_{.,t,k}}{\sigma_{Y_{.,t,k}}} \\ I_2 &= \sum_k \alpha_k \frac{Y_{i,t,k}}{Y_{.,t,k}} \\ I_3 &= \sum_k \alpha_k \frac{Y_{i,t,k}}{\max_i(Y_{i,t,k})} \end{aligned}$$

where  $Y_{.,t,k}$  is the mean of attribute  $k$  at date  $t$  and  $\sigma_{Y_{.,t,k}}$  its standard error. The first and the second indicators are centered on respectively zero and one at each period, and thus should be taken as relative human development indicators that provide a measure around the mean world human development level. The third one normalizes each attribute by its maximum value across countries, so that at each date  $t$  the mean value of the index depends on the distance between the leader country in a given attribute and the world mean value.

As a result, we obtained comparable weights between attributes, which means that the decomposition of the variance is robust to the scaling method. For the different indices the weights are respectively

- GDP per capita: 0.333, 0.327, 0.324
- Mean Human Capital: 0.345, 0.345, 0.354
- Life expectancy: 0.322, 0.328, 0.323

Figures 2-3-4 describe respectively indices 1, 2 and 3. Each curve represents the human development index of the countries included in a block e.g. Latin America or another. This index is based on three components for each country which belongs to the list of the 33 « countries » (this word refers in fact to one country, e.g. Germany or France, or to a group of countries, e.g. Scandinavian

countries). Therefore the internal distribution of income or human capital is not taken into account. These indices of human development are based on only  $33 \times 3 = 99$  figures.

Since the blocks' evolution for indices 1 and 2 are rather similar, we comment them together (see Figures 2 and 3). The most impressive performances since 1960 are the progress of Japan-Korea-Taiwan, China and other Asian countries. The first group of the former three countries, which took off before, is in the lead and has nearly reached Western Europe level, with the same average human capital, life expectancy, and for Japan the same income. The success of Eastern Europe in 1930-1960 is linked to rapid progress in education and longevity, thanks to the diffusion of primary school and health services by the socialist regimes. But since 1960 slow economic growth and stagnation of life expectancy in USSR explain a negative evolution.

The lowest values for index 2 are observed in South Asia and Africa in 2000 as well as in 1870. A very small improvement since 1960 has only compensated for a decline before. Between 1950 and 2000, the gap in income between Africa and the world average has increased, but the ratio average human capital in Africa/world average has rapidly increased from 0.24 to 0.63. This advantage explains a small improvement for the human development index despite the African reputation of economic failure. At the opposite, the leader block, Western Europe and offshoots, reveals a decline since 1930 which can surprise. It is a relative decline : average human capital and average longevity increase steadily in this block, but less than the world average.

Index 3 (figure 4) follows different trends because it is rescaled on the world leader country. We observe significant increases in each block since 1930, which proves a convergence between the leader country and the rest of the world. This convergence appears even among Western Europe countries since the index increases since 1870, with the exception of 1930-1950, as a consequence of World War II. The world convergence since 1930 has followed a phase of divergence : the gap between the leader country and Africa, South Asia, China and other Asian countries increased from 1870 to 1930. Such divergence results from the more rapid growth and progress in education in the United Kingdom and in the US, the first and second leaders.

The convergence starting in 1930 could be partially explained by the characteristics of the human development index. This index is weighted for 2/3 by mean human capital and life expectancy. A convergence process is predictable for these two components. The mean human capital in a block tends to a limit around 18 years and the progress of longevity are slower if the average life expectancy exceeds 80 instead of 50.

Then we computed inequality indices on the different indices. As it is impossible to use classic inequality indices on variables that can take non-positive values, we took the standard error of the various human development indices as a measure of inequality. Figure 5 shows an unambiguous increase of inequality until 1930 and a decline since that date for all indices. First, we must remember that internal inequality is not taken into account because these inequality indices are based only on the average values for 33 « countries ». From Table 1a

we know that internal inequality has slightly decreased for income distribution and much more for human capital distribution. So the inequality in human development between world citizens has been reduced rather more since 1930 than inequality between the 33 « countries ».

These two trends, rise until 1930 and decline after this year, are the consequences of increasing inequality for income as well as for longevity (the Theil has doubled) between 1870 and 1930, and of decreasing inequality for human capital (less 50%) and longevity (less 80%) since 1930. As income inequality is nearly the same in 2000 than in 1930, an increase of inequality in human development is impossible, even if the weights of human capital and life expectancy were lower.

But the choice of a rescaling method can have an important effect on the U-inverted curve of the inequality index. With the first index the variations are largely softened by the normalization of all variables, whereas the two other indicators induced important variations. These differences are not related to the weights of the attributes since the principal component analysis has given nearly the same weights.

The period between 1870 and 1930 is characterized by an increasing gap in income, life expectancy and education between Western Europe and the rest of the world, particularly Asia and Africa. There are some important interactions between the three components of human development : the success of industrialization and exports of manufactured goods in Western Europe results partially from the technological pace and the gains in productivity induced by important investment in education. The longevity could increase rapidly because average consumption of food improved and health services progressed thanks to education improvement. So that Western Europe populations were involved in a virtuous circle of human development which was out of reach of most other countries before 1950, even if small minorities had access to the same income or education.

But since 1930 and especially since 1950-1960, several other countries have progressively combined accumulation of human capital and increasing income. As a result they have obtained longer longevity and higher human development, whereas in Western Europe the rate of growth for average human capital and longevity was decreasing for the reasons given above i.e. a limit to years of schooling and difficulties to obtain a 10 % increase of life expectancy when the average exceeds 80.

## 6 Conclusion

This article presents an estimate of human development inequality on a long period, 1870-2000. For the first time a multidimensional assessment of world inequality which takes into account the main components of human development, income, education and longevity, is provided. Such an assessment was before impossible in the absence of human capital estimates since 1870. The standard error of different human development indices follows different evolu-

tions according to the index ; but we observe always an U-inverted curve with a maximum in 1930. Such a result is not surprising as longevity and income inequality have increased from 1870 to 1930, whereas human capital inequality and longevity inequality have largely fallen between 1930 and 2000, income inequality being nearly constant. Even with different weights for the three components of the human development index, a decrease of this index is certain since this trend depends on two falling components and a stable one. Except if we think that longevity and education do not matter, we can conclude that human development inequality has significantly decreased since 1930.

The availability of human capital estimates allows a comparison between the two globalization processes, the first in 1860-1914, the second since 1960 and mainly since 1980. In the world competition, human capital is a decisive advantage, particularly as the condition of access to technical progress. The situation has completely changed in a century. In 1910, 75% of individuals who achieved secondary schooling, lived in Western Europe and offshoots. In 2000, among people who have university education, the share of Western Europe is only 33%, whereas 42% live in Asia (including Japan, Korea and Taiwan). So there is a sort of discrepancy between the advantage of Western Europe and offshoots in world income distribution and the weight of Asia in world human capital.

Since 1980, the gathering pace of globalization coincides with decreasing world income inequality. So we observe since 1980 a triple convergence for income, longevity and human capital. If we refer to the leader country, the gap in human development between this country and other (except some African ones) is decreasing. Such convergence confirms, in a larger perspective, the prediction of R. Lucas (2000) for income distribution only. He predicted that the world income inequality would decrease during the 21st century, thanks to decreasing gaps between average national incomes. It is sure that the reduction of human development inequality will contribute to this process and that the decrease of human development inequality will be even more important than the decrease of income inequality.

## References

- Araujo C., F. Ferreira and N. Schady (2005). Is the world becoming more unequal? Changes in the World Distribution of Schooling. World Bank mimeo, forthcoming.
- Atkinson A and F.Bourguignon (1982).The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies*. P. 183-201.
- Barro R. and J.Lee (1997). International measures of schooling years and schooling quality. *American Economic Review*. P. 218-223.
- Barro R and J.Lee (2001). International data on educational attainment : updates and implications. *Oxford Economic Papers*. P.541-563.
- Becker G., T. Philipson and R. Soares (2005). The Quantity and Quality of Life and the Evolution of World Inequality. *American Economic Review*. P. 277–291.
- Berthélemy J-C. (2004). To what extent are African education policies pro-poor ?. Paris, mimeo.
- Bhalla S. (2002 ) Imagine there's no country : globalization and its consequences for poverty, inequality and growth.
- Bourguignon F. and C.Morrisson (2002). Inequality among world citizens : 1820-1992. *American Economic Review*. P. 727-744.
- Cohen D. and M.Soto (2001). Growth and human capital : good data, good results. Technical Papers n° 179. Paris.OECD Development Centre.
- De La Fuente A. and R.Domenech (2001). Educational attainment in the OECD : 1960-1995. Mimeo.
- Firebaugh G. (2003). The new geography of global income inequality. Cambridge ; Harvard U.P.
- Lucas R. (2000). Some macro-economics for the twenty-first century. *Journal of Economic Perspectives*. P.159-168.
- Lucas R. (2002). Lectures on economic growth. Cambridge. Harvard U.P.
- Maasoumi, E. (1989). Composite Indices of Income and Other Developmental Indicators: A General Approach. *Research on Economic Inequality*, Vol.1, 269-286.
- Maddison A. (2003). The world economy. Historical statistics. Paris.OECD Development Centre.
- Nehru V., E. Swanson and A.Dubey (1995). A new database on human capital stocks in developing and industrial countries : sources, methodology and results. *Journal of Development Economics*. P.379-401.
- Ram, R. (1982). Composite Indices of Physical Quality of Life, Basic Needs-Fulfillment, and Income: A 'Principal Component' Analysis. *Journal of Development Economics*, 11, 227-247.
- Sala-I-Martin X. (2002). The world distribution of income. New-York. Columbia University. Mimeo.
- Sevedberg P. (2004). World income distribution : which way ?. *The Journal of Development Studies*. P.1-32.
- Vinod T. Y.Wang and X.Fan (2001). Measuring education inequality : Gini coefficients of education for 140 countries, 1960-2000. World Bank, mimeo.

World Bank (2005). Equity and Development, World Development Report 2006.

## A Tables

Table 1 a - The World Distribution of Income, Education and Life Expectancy

	1870	1890	1910	1929	1950	1960	1970	1980	1990	2000
<b>Income<sup>1</sup></b>										
Coefficient of Gini	0.553	0.586	0.614	0.635	0.648	0.644	0.652	0.665	0.664	0.646
Theil Index	0.669	0.762	0.833	0.813	0.775	0.788	0.780	0.854	0.853	0.876
within country groups	0.481	0.511	0.509	0.410	0.293	0.303	0.287	0.325	0.325	0.382
between country groups	0.188	0.251	0.324	0.403	0.482	0.485	0.492	0.529	0.528	0.494
Mean Logarithmic Deviation	0.527	0.605	0.678	0.749	0.800	0.798	0.825	0.883	0.852	0.793
within country groups	0.339	0.354	0.354	0.346	0.318	0.313	0.332	0.354	0.324	0.299
between country groups	0.162	0.217	0.299	0.381	0.471	0.508	0.518	0.577	0.517	0.472
Mean Income (PPP \$ 1990)	890	1113	1453	1768	2145	2759	3774	4483	4922	6035
<b>Years of Schooling</b>										
Coefficient of Gini	0.793	0.755	0.709	0.654	0.601	0.568	0.533	0.486	0.438	0.394
Theil Index	1.432	1.248	1.074	0.902	0.762	0.668	0.577	0.477	0.388	0.321
within country groups	0.842	0.667	0.555	0.473	0.411	0.381	0.355	0.305	0.259	0.234
between country groups	0.590	0.581	0.519	0.429	0.351	0.287	0.222	0.172	0.129	0.087
Mean Logarithmic Deviation	3.184	3.171	3.060	2.826	2.592	2.401	2.129	1.820	1.523	1.286
within country groups	2.594	2.590	2.541	2.397	2.241	2.114	1.907	1.648	1.394	1.199
between country groups	0.933	0.762	0.615	0.485	0.385	0.310	0.230	0.174	0.130	0.087
Mean Years of Schooling	1.06	1.50	2.00	2.53	3.26	3.90	4.63	5.25	6.00	6.70
Illiteracy Rate	74.5	69.4	63.6	56.2	49.4	44.3	38.0	31.6	25.8	21.7
<b>Life Expectancy<sup>1</sup></b>										
Mean Life Expectancy	28.4	29.9	32.8	38.5	50.1	55.7	59.4	60.6	61.1	66.3
Theil Index (between countries)	0.0237	0.032	0.045	0.046	0.025	0.016	0.012	0.012	0.013	0.009
Population (millions)	1267	1451	1722	2044	2507	3021	3663	4419	5314	6071

<sup>1</sup>source of income distributions and life expectancy: Bourguignon-Morrisson (2002)

source of mean income: Maddison (2003)



Table 1 b - The World Distribution of Income, Education and Life Expectancy

	1870	1890	1910	1929	1950	1960	1970	1980	1990	2000
<b>Income<sup>1</sup></b>										
Coefficient of Gini	0.553	0.586	0.614	0.635	0.648	0.644	0.652	0.665	0.664	0.646
Theil Index	0.669	0.762	0.833	0.813	0.775	0.788	0.780	0.854	0.853	0.876
within country groups	0.481	0.511	0.509	0.410	0.293	0.303	0.287	0.325	0.325	0.382
between country groups	0.188	0.251	0.324	0.403	0.482	0.485	0.492	0.529	0.528	0.494
Mean Logarithmic Deviation	0.527	0.605	0.678	0.749	0.800	0.798	0.825	0.883	0.852	0.793
within country groups	0.339	0.354	0.354	0.346	0.318	0.313	0.332	0.354	0.324	0.299
between country groups	0.162	0.217	0.299	0.381	0.471	0.508	0.518	0.577	0.517	0.472
Mean Income (PPP \$ 1990)	890	1113	14533	1768	2145	2759	3774	4483	4922	6035
<b>Years of Schooling</b>										
Coefficient of Gini	0.835	0.810	0.778	0.737	0.695	0.663	0.623	0.579	0.534	0.497
Theil Index	1.575	1.410	1.248	1.079	0.936	0.839	0.727	0.621	0.524	0.454
within country groups	0.985	0.829	0.729	0.650	0.585	0.552	0.505	0.449	0.395	0.367
between country groups	0.590	0.581	0.519	0.429	0.351	0.287	0.222	0.172	0.129	0.087
Mean Logarithmic Deviation	2.590	2.617	2.561	2.418	2.267	2.134	1.936	1.708	1.467	1.274
within country groups	1.657	1.855	1.946	1.933	1.882	1.824	1.706	1.534	1.337	1.187
between country groups	0.933	0.762	0.615	0.485	0.385	0.310	0.230	0.174	0.130	0.087
Mean Years of Schooling	1.04	1.44	1.88	2.39	3.09	3.70	4.57	5.56	6.39	7.03
Illiteracy Rate	74.5	69.4	63.6	56.2	49.4	44.3	38.0	31.6	25.8	21.7
<b>Life Expectancy<sup>1</sup></b>										
Mean Life Expectancy	28.4	29.9	32.8	38.5	50.1	55.7	59.4	60.6	61.1	66.3
Theil Index (between countries)	0.0237	0.032	0.045	0.046	0.025	0.016	0.012	0.012	0.013	0.009
Population (millions)	1267	1451	1722	2044	2507	3021	3663	4419	5314	6071

Table 2 - Education shares (in percent)

	<b>1870</b>	<b>1890</b>	<b>1910</b>	<b>1930</b>	<b>1950</b>	<b>1960</b>	<b>1970</b>	<b>1980</b>	<b>1990</b>	<b>2000</b>
<b>Bottom 60</b>	0.6	0.4	0.3	3.7	12.3	16.7	20.8	23.8	28.1	31.6
<b>Bottom 80</b>	10.9	21.7	31.3	39.0	45.0	48.7	49.9	53.9	57.1	60.1
<b>Top 10</b>	52.7	47.2	41.8	36.0	32.9	31.2	30.1	26.0	24.0	22.0
<b>Top 5</b>	32.1	27.1	22.3	20.0	19.7	18.7	16.5	14.1	13.2	12.0

Table 3 - Regional Composition of World Education Quantiles

	Africa	Asia	Japan Korea Taiwan	Latin America	Eastern Europe	Europe and offshoots	Total
<b>1870</b>							
Total	6.6	57.1	3.2	2.7	10.4	18.2	100
Bottom 80	7.7	63.0	3.4	2.9	10.6	12.4	100
Decile 9	0.6	17.0	8.2	1.8	13.4	59.0	100
Decile 10	0.1	31.7	0.6	1.8	9.6	55.6	100
<b>1910</b>							
Total	6.3	51.3	3.5	3.7	12.7	21.7	100
Bottom 80	7.6	58.8	3.1	4.1	13.2	13.2	100
Decile 9	1.1	28.6	9.2	3.5	16.7	40.9	100
Decile 10	0.5	15.9	2.4	0.9	6.4	73.9	100
<b>1960</b>							
Total	9.1	50.2	4.3	6.1	10.8	19.3	100
Bottom 80	11.2	59.8	2.7	6.9	7.7	11.7	100
Decile 9	4.1	28.7	8.8	9.4	22.0	27.0	100
Decile 10	1.7	17.2	9.8	2.1	20.4	48.8	100
<b>1980</b>							
Total	10.6	53.3	3.9	7.2	9.3	15.7	100
Bottom 80	13.7	63.8	1.9	7.8	6.6	6.3	100
Decile 9	3.9	30.7	8.1	7.0	19.1	31.3	100
Decile 10	2.4	24.7	9.6	5.0	13.9	44.2	100
<b>2000</b>							
Total	13.2	55.1	3.2	7.8	7.6	13.2	100
Bottom 80	16.8	63.8	1.3	7.2	5.4	5.4	100
Decile 9	6.1	41.5	5.8	10.3	12.0	24.3	100
Decile 10	5.4	33.7	8.3	7.7	11.9	32.9	100

Calculations obtained after smoothing of the cumulative distribution function of education

Table 4 - Average Human Capital and Illiteracy Rates, 1870-2000

		Africa	Japan Korea Taiwan	Latin America	Eastern Europe	Europe and offshoots	Asia	China	South Asia	Other Asian countries
<b>1870</b>	<b>Total</b>	<b>0.12</b>	<b>0.61</b>	<b>0.67</b>	<b>0.98</b>	<b>3.16</b>	<b>0.55</b>	<b>1.00</b>	<b>0.03</b>	<b>0.19</b>
	Primary	0.10	0.59	0.61	0.88	3.02	0.50	0.91	0.02	0.17
	Illiteracy	96.3	72.6	84.9	76.5	30.2	86.9	79.0	97.0	90.7
<b>1910</b>	<b>Total</b>	<b>0.25</b>	<b>2.58</b>	<b>1.11</b>	<b>1.65</b>	<b>5.79</b>	<b>0.77</b>	<b>1.20</b>	<b>0.34</b>	<b>0.31</b>
	Primary	0.24	2.50	1.03	1.49	5.54	0.69	1.10	0.27	0.27
	Illiteracy	93.4	39.5	70.7	62.8	13.3	83.4	76.0	93.0	85.8
<b>1950</b>	<b>Total</b>	<b>0.78</b>	<b>6.28</b>	<b>2.57</b>	<b>4.91</b>	<b>7.64</b>	<b>1.34</b>	<b>1.60</b>	<b>1.00</b>	<b>1.41</b>
	Primary	0.71	4.85	2.43	4.44	6.73	1.16	1.44	0.76	1.29
	Illiteracy	85.1	9.5	41.3	15.0	5.0	75.2	72.0	81.6	69.7
<b>1970</b>	<b>Total</b>	<b>1.63</b>	<b>8.64</b>	<b>4.23</b>	<b>7.80</b>	<b>9.65</b>	<b>2.66</b>	<b>3.10</b>	<b>1.95</b>	<b>2.99</b>
	Primary	1.39	5.72	3.49	6.55	7.15	2.08	2.48	1.40	2.52
	Illiteracy	71.3	7.2	29.1	9.5	2.6	54.0	48.9	67.6	39.2
<b>1980</b>	<b>Total</b>	<b>2.26</b>	<b>9.90</b>	<b>5.03</b>	<b>8.17</b>	<b>10.73</b>	<b>3.47</b>	<b>4.10</b>	<b>2.52</b>	<b>3.90</b>
	Primary	1.79	6.20	3.63	6.49	6.84	2.51	2.91	1.78	3.06
	Illiteracy	61.2	4.2	21.5	7.6	2.1	43.0	34.0	61.5	26.6
<b>1990</b>	<b>Total</b>	<b>3.24</b>	<b>10.94</b>	<b>6.52</b>	<b>9.48</b>	<b>11.44</b>	<b>4.37</b>	<b>5.06</b>	<b>3.04</b>	<b>5.55</b>
	Primary	2.50	6.52	4.44	7.06	7.07	3.13	3.49	2.17	4.30
	Illiteracy	50.3	2.5	15.6	5.4	1.7	33.6	22.0	53.8	17.7
<b>2000</b>	<b>Total</b>	<b>4.19</b>	<b>11.79</b>	<b>7.42</b>	<b>9.66</b>	<b>11.78</b>	<b>5.34</b>	<b>5.96</b>	<b>4.12</b>	<b>6.69</b>
	Primary	3.24	7.00	5.08	7.17	7.24	3.85	4.11	2.96	5.19
	Illiteracy	40.0	1.6	12.5	3.3	1.4	27.2	15.0	47.3	12.4

## B Annexes

### B.1 Annex 1

This data set 1870-2000 on human capital is based on two data sets. The first for 1870-1960 is a new one, the second for 1960-2000 has been published by D.Cohen and F.Soto (2001) (it is quoted as Cohen-Soto). The latter has been completed because several countries (as Bangladesh, Poland, USSR/Russia) were missing on the list. We have given new estimates until 2000 for these countries. The Cohen-Soto data set has been chosen because it provides the most reliable estimates as proved by a comparison of these data with the data sets published by R. Barro and J-W Lee (2000), V.Nehru et al. (1995), A. De La Fuente and R.Domenech (2000). Cohen-Soto has systematically used as sources the national censuses which give the school attainments of population (usually aged 15 to 64). The long series of B.Mitchell (1998) on primary, secondary and high school enrolments were used only to fill missing cells.

The 1870-1960 data set is a new one for all countries. As censuses are scarce before 1960, except in the most developed countries like the U.S., we used all the available sources, like the long series of enrolment figures of Mitchell (1998), quoted as Mitchell, and censuses which give school attainments or literacy rates and other sources as the percentages of males and females unable to sign at marriage. We followed Cohen-Soto's methodology : When differences arise between two sources, censuses data are always preferred. For example, if a census gives a literacy rate of 50% and the Mitchell series an estimate of 1 average year of schooling, we conclude that the right figure is 2 or 2.5 years since 50% of population received 4 or 5 years of primary schooling. So this data set for 1870-2000 is consistent because the same methodology has been applied for 130 years. When our own estimate for 1960 is the same, or nearly the same, than Cohen-Soto's estimate, we kept his figure.

Estimating the average years of schooling in 1910 or before is difficult because information on school enrolments before 1870 are needed. Mitchell provides series for European countries, U.S., Canada, Australia before 1870, but in Latin America, in Eastern Europe, in some Asian countries, the series begin only around 1870 or 1880. Moreover for African countries, other Asian countries, Mitchell gives no data before 1930 or 1950-60. So we estimated the average years of schooling in all countries where series are not available by interpolation. We assumed an enrolment rate in primary school of 1% in 1820 and a constant rate of increase between 1820 and the first year of Mitchell's series. Some simulations prove that we obtain nearly similar results if we assumed an enrolment rate of 3 or 5% in 1820.

- ASIA.

Philippines. 1870-1960 : Mitchell ; 1970-2000 : Cohen-Soto

Thailand. 1870-1910 : Mitchell ; 1920-1960 : Series of Mitchell adjusted on the basis of literacy rates in 1937, 1950, 1960 and of the census in 1947

which gives school attainments (Unesco 1957, 1965), 1970-2000 : Cohen-Soto's estimates adjusted on the basis of the literacy rate in 1980.

Indonesia. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1961 (Unesco 1957, 1965), 1960-2000 : Cohen-Soto.

Turkey. 1870-1920 : from A.Kazamias (1966), 1930-1950 : series of Mitchell adjusted on the basis of literacy rates in 1935, 1950 (Unesco 1957), 1960-2000 : Cohen-Soto.

China. 1870-1920 : from E. Sakakida Rawski (1979), 1930-1950 : from A.Maddison (1998) and statistics given by M. Bastid-Bruguière. 1960-2000 : Cohen-Soto.

Japan. 1870-2000 : series of Y.Godo and Y.Hayami (2002).

India. 1870-1960 : Mitchell, 1970-2000 : Cohen-Soto.

Bangladesh. 1870-1950 : data of India ; 1960 : interpolation between 1950 and 1970 ; 1970-2000 : Cohen-Soto.

Pakistan. 1870-1950 : data of India ; 1960 : interpolation between 1950 and 1970 ; 1970-2000 : R.Barro and J.W.Lee (2000).

Myanmar. 1870-1960 : Mitchell, 1970-2000 : Cohen-Soto.

Korea. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in 1930, 1950 (Unesco 1957), 1960-2000: Cohen-Soto.

Taiwan. 1870-1960 : series of Mitchell adjusted on the basis of the data of China and of the literacy rates in 1950, 1956 and the enrollment rates in 1935 (Unesco 1957, 1965) ; 1970-2000 : R.Barro and J.W.Lee (2000).

Other countries

Irak. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950 and 1957 (Unesco 1957, 1965), 1970-2000 : Cohen-Soto.

Iran. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1956 and 1966 (Unesco 1957, 1965, 1970), 1970-2000 : Cohen-Soto.

Malaysia. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1930, 1950 and 1960 (Unesco 1957, 1965) ; 1970-2000 : Cohen-Soto.

Sri-Lanka. 1870-1960 : Mitchell ; 1970 : interpolation between 1960 and 1980 ; 1980-2000 : R.Barro and J.W.Lee (2000).

Syria. 1870-1960 : Mitchell ; 1970-2000: Cohen-Soto.

Vietnam. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1960 (Unesco 1957, 1965), 1970-1980: interpolation between 1960 and 1990. 1990: R.Barro and J.W. Lee (2000) . 2000: Extrapolation.

#### • NORTH AND SOUTH AMERICA.

Canada. 1870 ; series of Mitchell adjusted on the basis of data in other developed countries. 1880-1900 : interpolation between 1870 and 1910. 1910-1950 : Mitchell. 1960-2000 : Cohen-Soto.

United States. 1870-1950 : Mitchell. 1960-2000 : Cohen-Soto.

Mexico. 1870-1950 : series of Mitchell adjusted on the basis of national censuses (S.Hunt 2002), 1960 : interpolation between 1950 and 1970. 1970-2000 : Cohen-Soto.

Argentina, Brazil, Chile, Colombia, Peru, Venezuela : 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in several years (S.Engerman et al. 2000).

Other countries

Bolivia, Cuba, Guatemala, Jamaica, Nicaragua, Paraguay, Salvador : 1870-1950 : Mitchell ; 1960-2000 : Cohen-Soto.

Costa-Rica. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1892,1925 (S.Engerman et al. 2000) . 1970-2000 : Cohen-Soto.

Dominican Republic 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1956 (Unesco 1965) ; 1970-2000 : Cohen-Soto.

Ecuador. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950,1962 (Unesco 1965).

1970-2000 : Cohen-Soto.

Honduras. 1870-1950 : series of Mitchell adjusted on the basis of the literacy in several years (S.Engerman et al. 2000), 1960-2000: Cohen-Soto.

Panama. 1870 : figure of Uruguay in 1870 instead of Mitchell estimate which seems debatable

1880-1930 : interpolation between 1870 and 1940 ; 1940-1960 : Mitchell. 1970-2000 : Cohen-Soto.

Uruguay. 1870-1950 :series of Mitchell adjusted on the basis of the literacy rates in several years (S.Engerman et al. 2000, Unesco 1957) ; 1960 : interpolation between 1950 and 1970 ; 1970-2000 : Cohen-Soto.

#### • AFRICA

Algeria, Tunisia : 1870-1950 : Mitchell ; 1960-2000 : Cohen-Soto.

Morocco : 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in 1950 ( Unesco 1957), 1960-2000 : Cohen-Soto.

Egypt. 1870-1930 : Mitchell ; 1950, 1960 : series of Mitchell adjusted on the basis of the literacy rates in 1947, 1960 (Unesco 1965) ; 1970-2000: Cohen-Soto.

Côte-d'Ivoire, South Africa. 1870-1950 : Mitchell. 1960-2000 : Cohen-Soto.

Ghana, Kenya and Nigeria. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rate in 1950 (Unesco 1957) ; 1960-2000 : Cohen-Soto.

Other countries.

Angola, Cameroon, Mozambique, Zimbabwe. 1870-1950 : Mitchell. 1960-2000 : Cohen-Soto.

Zaire. 1870-1950 : Mitchell. 1960-2000 : R.Barro and J.W. Lee (2000).

Benin. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rate in 1950 (Unesco 1957). 1960-2000 : Cohen-Soto.

Ethiopia. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rate in 1950 (Unesco 1957) ; 1960-1980 : Cohen-Soto estimates adjusted on the basis of the literacy rate in 1980 (Unesco 1999 ) ; 1990 and 2000 : Cohen-Soto.

Madagascar and Malawi. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in several years (Unesco 1957, 1965), 1970 : interpolation between 1960 and 1980 . 1980-2000 : Cohen-Soto.

Niger. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1960 (Unesco 1957, 1965). 1970-2000 : Cohen-Soto.

Senegal. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rate in 1961 (Unesco 1965) ; 1970-2000 : Cohen-Soto.

Sudan and Tanzania. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in several years (Unesco 1957, 1965) ; 1970-2000 : Cohen-Soto.

Zambia. 1870-1940 : Mitchell ; 1950 and 1960 : series of Mitchell adjusted on the basis of the literacy rates in 1950, 1963 (Unesco 1957, 1965) ; 1970-2000 : Cohen-Soto.

- EUROPE.

Ireland and United-Kingdom.

Ireland. 1870-1950 : Mitchell ; 1960-2000 : Cohen-Soto.

United-Kingdom : 1870-1950 : series of Mitchell adjusted on the basis of H.Graff ( ). 1960-2000 : Cohen-Soto.

Portugal and Spain.

Portugal. 1870-1950 : Mitchell. 1960-2000 : Cohen-Soto.

Spain. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rate in 1910 (Unesco 1957) and data in Johanson (1977). 1970-2000 : Cohen-Soto.

Germany. 1870-1950 : Mitchell ; 1960-2000 : Cohen-Soto.

France. 1870-1960 : series of Mitchell adjusted on the basis of the enrolment in primary schools (INSEE, 1966). 1970-2000: Cohen-Soto.

Benelux and Switzerland.

Belgium. 1870-1960 : Mitchell. 1970-2000 : Cohen-Soto.

Netherland and Switzerland. 1870-1960 : series of Mitchell adjusted on the basis of the enrolment in secondary schools (P.Flora et al. 1983). 1970-2000 : Cohen-Soto.

Scandinavian countries.

Denmark. 1870-1960 : Mitchell ; 1970-2000 : Cohen-Soto.

Finland. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rate in 1930 (Unesco 1957). 1970-2000: Cohen-Soto.

Norway and Sweden. 1870-1960 : series of Mitchell adjusted on the basis of the enrolment in secondary schools (P.Flora et al. 1983). 1970-2000: Cohen-Soto.

Austria, Czechoslovakia and Hungary.

Austria and Hungary. 1870-1960 : Mitchell. 1970-2000 : Cohen-Soto.

Czechoslovakia. 1870-1950 : the same enrolment rates that in Austria (cf. the same literacy rates in H.Graff). 1960-2000: R.Barro and J.W.Lee (2000).

Poland.1870-1950 : Mitchell. 1960-2000 : R.Barro and J.W.Lee (2000).

USSR/Russia. 1870-1950 : estimates on the basis of Ben Eklof (1986), (H.Graff) and (Unesco 1957). 1960-2000 : R.Barro and J.W.Lee (2000).

Bulgaria, Greece, Romania and Yugoslavia.

Bulgaria. 1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1900, 1910, 1930 (Unesco 1957). 1970-2000: Cohen-Soto.



Greece. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in 1910, 1928 and 1950 (Unesco 1957). 1960-2000: Cohen-Soto.

Romania.1870-1960 : series of Mitchell adjusted on the basis of the literacy rates in 1930, 1948 and 1956 (Unesco 1957, 1965).1970-2000 : Cohen-Soto.

Yugoslavia. 1870-1950 : series of Mitchell adjusted on the basis of the literacy rates in 1921, 1931, 1948 and the census of 1953 which gives the school attainment of the population (Unesco 1957). 1960-2000 : R.Barro and J.W.Lee (2000).

- OCEANIA

Australia. 1870-1960 : Mitchell ; 1970-2000 : Cohen-Soto.

New-Zealand : estimates of Australia.

## References for Annex 1

R.Barro and J.W.Lee (2000) . International Data on Educational Attainment : Updates and Implications. CID Working Paper n°42, Harvard University.

Ben Ekloff (1986) Russia Peasants Schools. University of California Press , Berkeley.

D.Cohen and M.Soto (2001). Growth and Human Capital : Good Data, Good Results. Technical paper n° 179 . Development Centre, OECD, Paris.

A.De La Fuente and R Domenech (2001).Educational Attainment in the OECD, 1960-1995 , manuscript.

S.Engerman, S.Haber and K. Sokoloff (2000). Inequality, Institutions and Differential Paths of Growth Among New World Economy ; in Institutions, Contracts and Organizations. C.Menard ed: Edward Elgar. Cheltenham.

P.Flora et al. (1983). State, Economy and Society in Western Europe, 1815-1975. Frankfurt : Campus Verlag.

Y Godo and Y Hayami (2002). « Catching up in Education in the Economic Catch-Up of Japan with the United States, 1890-1990 ». Economic Development and Cultural Change. P. 961-978.

H.Graff ( 1987) The Legacies of Literacy. Indiana U.P. Blooming.

S.Hunt (2002). Catching up in Education During the Twentieth Century : the U.S. and Latin America. Mimeo.

INSEE (1966). Annuaire Statistique de la France, résumé rétrospectif. 1966. Paris.

Johanson (1977). The History of Literacy in Sweden. Umea.

A.Kazamias (1966) Education and the Quest for Modernity in Turkey. University of Chicago Press, Chicago.

P.Lindert (2004) Growing Public. Vol. 2. Cambridge U.P. Cambridge.

A.Maddison (1998) Chinese Economic Performance in the Long Run .Development Centre, OECD,Paris.

B.Mitchell (1998) International Historical Statistics : the Americas 1750-1993. M.Stockton Press, New-York.

B.Mitchell (1998) International Historical Statistics : Europe 1750-1993. M.Stockton Press, New-York.

B.Mitchell (1998) International Historical Statistics : Africa, Asia and Oceania 1750-1993, M. Stockton Press, New-York.

V.Nehru, E.Swanson and A.Dubey (1995) « A New Data Base on Human Capital Stocks in Developing and Industrial Countries: Sources, Methodology and results » Journal of Development Economics, p. 379-401.

E. Sakakido-Rawski (1979) Education and Popular Literacy in Ch'ing China. University of Michigan Press, Ann Arbor.

UNESCO (1957) World Illiteracy at Mid-Century. Paris.

UNESCO (1965, 1970 , 1999) Statistical Yearbook, Paris.

## B.2 Annex 2

### B.2.1 Assumptions underlying Mitchell's series

At a given period, the educational situation of a country can be assessed directly by census data, provided that it exists, or can be derived from demographic and educational information over the past generations. The latter procedure estimates the mean educational attainment of cohorts of  $i$  years-old individuals at date  $t$  by computing the enrollment rate in the primary school at date  $t-i+6$ , and by relying on an estimate of duration at school. Such a procedure introduces much uncertainty than census data, but enables us to recover educational data over very long periods for which census data does not exist.

Some problems have been recognized to arise from this enrollment-based procedure. First, the population's structure in year  $t$  is not necessarily the outcome of year  $t-T$  given a mortality rule between those two periods, because migrations can affect a substantial proportion of population. Between the 19th and the 20th century, countries from the Commonwealth, Latin America, North-America, and some of Europe have had intense periods of migrations. Depending on the human capital of the migrants relatively to their compatriots, the net impact of migration can be positive or negative. A second problem is that the intake rate, i.e. the ratio of new entrants in primary school to the six-years population, is subject to measurement errors due to the presence of repeaters and dropouts. We derive human capital measurement by ignoring the migration problem; we believe that corrections based on literacy rates reduce this problem in some extent.

Let  $P_{i,t}$  be the population of age  $i$  at time  $t$ ,  $E_t$  and  $N_t$  be respectively the total number of pupils at school and the number of intakes - those attending their first year of school in year  $t$ . Given a cohort of age  $i$  at time  $t$ , the probability to have been an intake at the age of 6 is simply

$$\frac{N_{t-i+6}}{P_{6,t-i+6}} \quad (1)$$

As in Cohen and Soto (2001) we consider the impact of repeaters and dropouts by assuming that a pupil can repeat a maximum of three years during her scolarity, which lasts  $P$  years. Let  $d$  and  $r$  be the dropout and repeating rates, and  $g$  the growth rate of intakes. The expression linking total enrollment  $E_t$  to first-year enrollment  $N_t$  is

$$\begin{aligned} E_t &= N_t \sum_{j=0}^{P-1} (1-d-r)^j \left[ \frac{r}{(1+g)^j} + \frac{r \binom{j+1}{1}}{(1+g)^{j+1}} + \frac{r^2 \binom{j+1}{2}}{(1+g)^{j+2}} + \frac{r^3 \binom{j+1}{3}}{(1+g)^{j+3}} \right] \\ &= N_t \mu(d, r, g, P) \end{aligned} \quad (2)$$

This formula simply decomposes each grade at school between students who have repeated 0, 1, 2 or 3 times before. Our data provides total enrollment  $E_t$ , from which is deduced the number of intakes  $N_t$  from 1870 to 1960. Then a

cohort  $i$  at time  $t$  displays a mean number of schooling equal to

$$\frac{N_{t-i+6}}{P_{6, t-i+6}} \cdot \left( \sum_{j=0}^{P-1} j(1-d)^j \cdot d + P(1-d)^P \right) = \frac{N_{t-i+6}}{P_{6, t-i+6}} \lambda(d, P) \quad (3)$$

In this equation the  $\lambda(d, P)$  term is the mean duration of primary school which is held constant over time and does not take into account repeated years. From (2) and (3), human capital  $H_{i, t}$  of cohort  $i$  at time  $t$  is given by

$$H_{i, t} = \frac{E_{t-i+6}}{P_{6, t-i+6}} \frac{\lambda(d, P)}{\mu(d, r, g, P)} \quad (4)$$

In the case where  $d = r = g = 0$ , one simply has  $H_{i, t} = E_{t-i+6}/P_{6, t-i+6}$  since  $\lambda(d, P) = \mu(d, r, g, P) = P$ . Furthermore human capital does not depend on any assumption on the duration  $P$  of schooling, since there is a perfect trade-off between the mean number of years at school ( $\lambda$ ) and the mean number of pupils at each grade ( $E/\mu$ ).

The data consists in demographic and enrollment files beginning in various years. The demographic files present the structure of the population by age group. As we need the population of age  $i$  in year  $t$ , we first smooth the data by assuming that the distribution of mortality  $F$  is Weibull  $(a, b)$ , which parameters are calibrated on the life expectancy of the population and the survival rate after 60 years (taken equal to 10%). Life expectancy is corrected from children mortality, equal to  $m_0 = 20\%$  at birth and to  $m_1 = 7\%$  the following 4 years. Formally life expectancy  $LE$  is given by

$$\begin{aligned} LE &= m_0 + m_1(2 + 3 + 4 + 5) + (1 - m_0)(1 - m_1)^4 \sum_{k \geq 6} p_k k, \quad p_k \rightsquigarrow \text{Weibull}(a, b) \\ &= \nu(m_0, m_1, a, b) \end{aligned} \quad (5)$$

Once calibrated, the survival function  $1 - F$  gives the relative weight of each cohort of age inside each age group

$$\frac{p(\text{Age} = i)}{p(\text{Age} = j)} = \frac{1 - F(\text{Death} \leq i)}{1 - F(\text{Death} \leq j)} \quad (6)$$

Dropout and repetition rates are derived from Unesco data in 1960. When data is missing, it is assumed that the demographic structure is the same in 1820 and in 1850 than in 1872, as well as the enrollment rate at school. Last, data is smoothed across years using spline functions.

## B.2.2 The smoothing procedure

Figure 1 describes easily the principle of smoothing that has been used: we computed an upper bound for the cumulative distribution function by interpolating the left corners of the empirical cumulative function (similar to a stair function), and a lower bound by interpolating the right corners. The mean of these bounds is the smoothed cumulative distribution function.

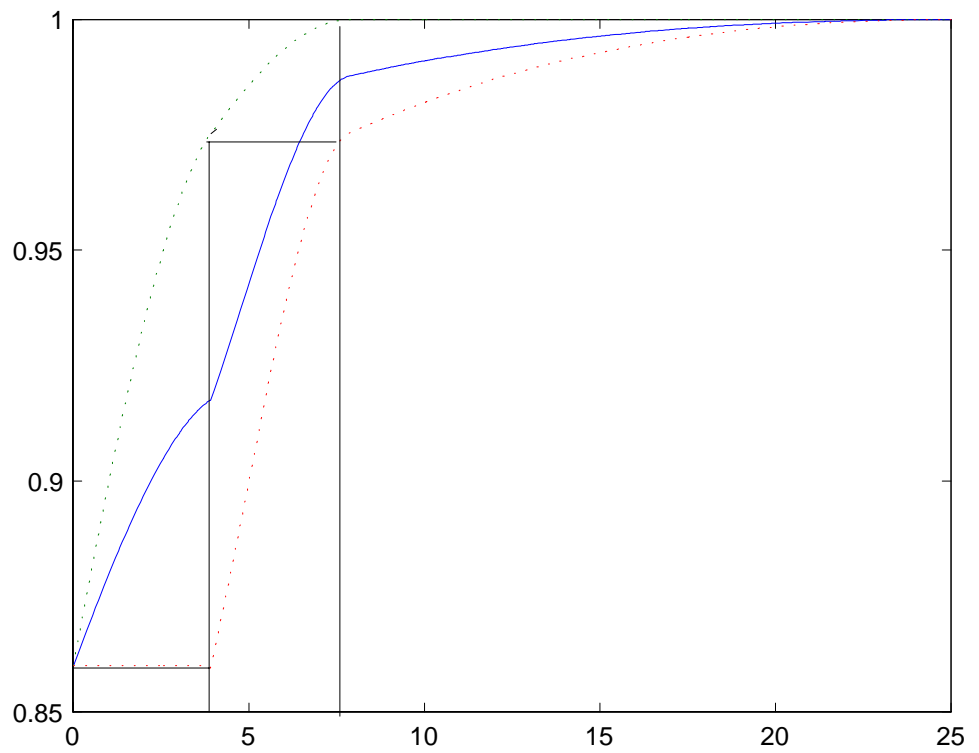


Figure 1: The three-groups education cumulative distribution and its smoothing.

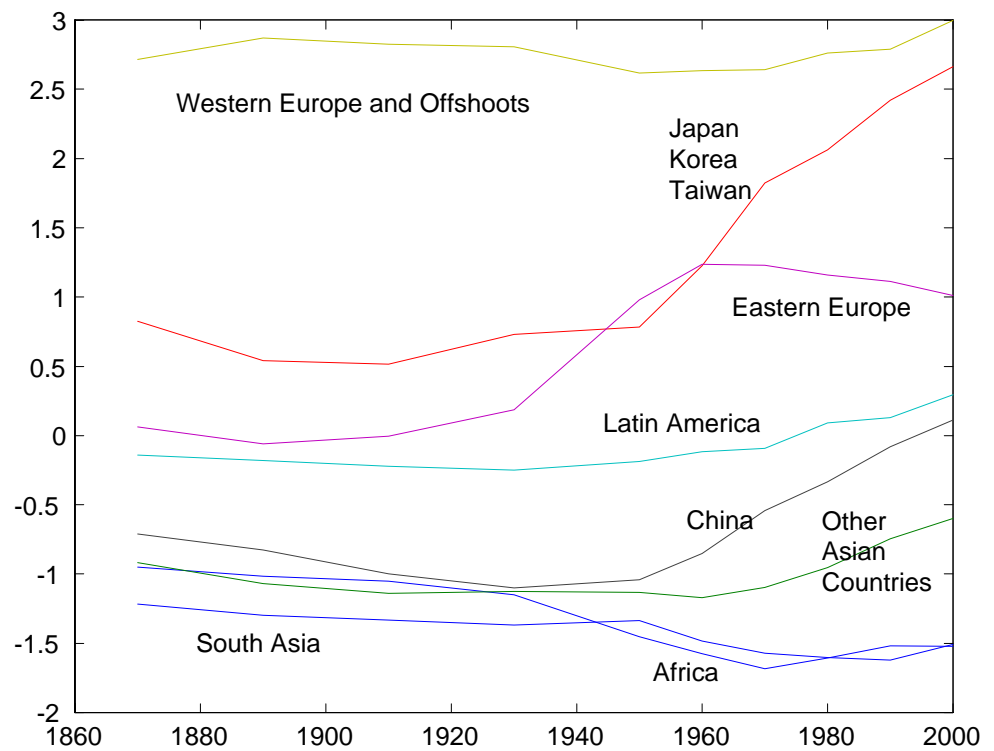


Figure 2: Evolution of Human Development for Some Groups of Countries - Index 1

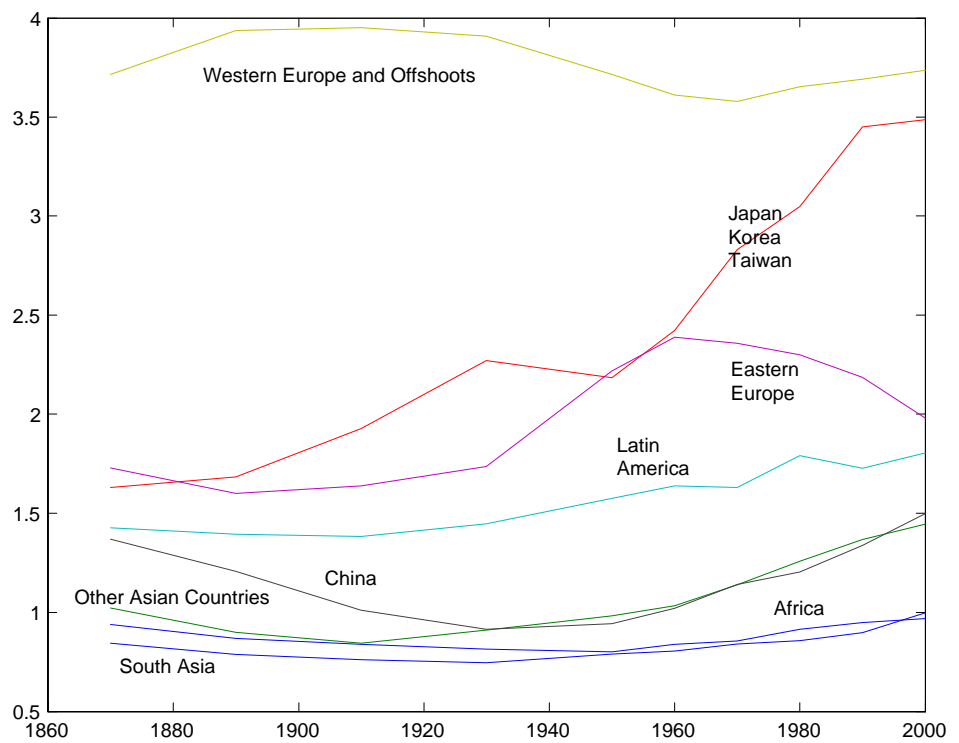


Figure 3: Evolution of Human Development for Some Groups of Countries - Index 2



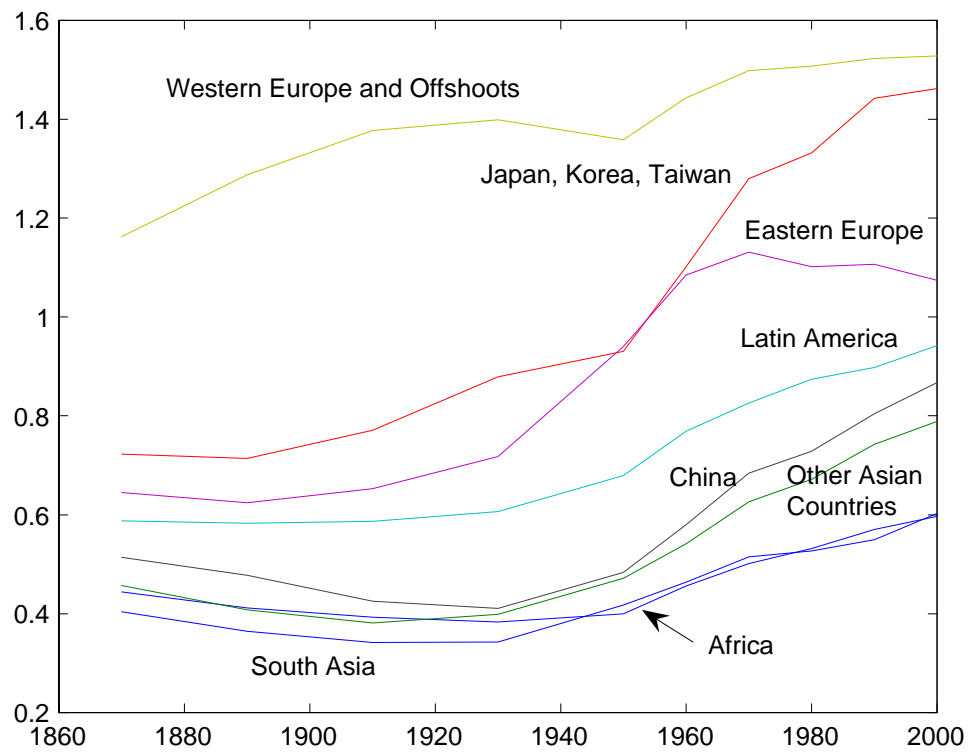


Figure 4: Evolution of Human Development for Some Groups of Countries - Index 3

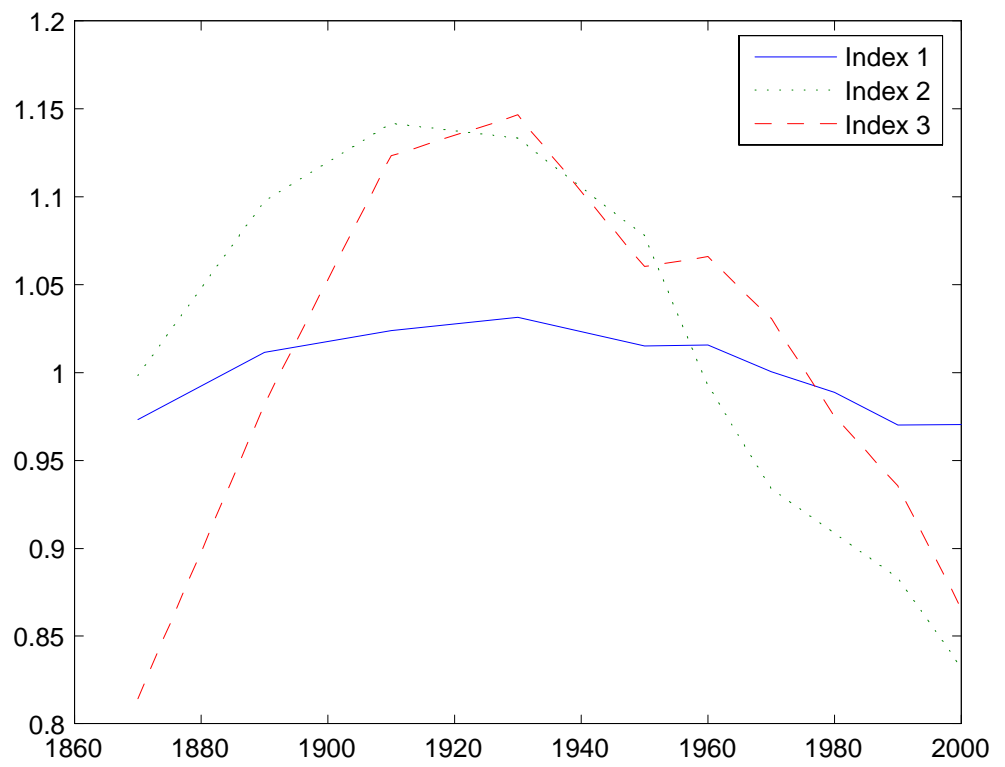


Figure 5: Inequality in Human Development