ISSN 1352-2469 Distributional Analysis Research Programme

Education, Inequality and Income Inequality

by

Daniele Checchi
University of Milan

Discussion Paper
No. DARP 52
May 2001

Distributional Analysis Research Programme
The Toyota Centre
Suntory and Toyota International
Centres for Economics and
Related Disciplines
London School of Economics
Houghton Street
London WC2A 2AE

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge and most papers are downloadable from the website. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Sue Coles on:

Telephone:	UK+20 7955 6678
Fax:	UK+20 7955 6951
Email:	s.coles@lse.ac.uk
Web site:	http://sticerd.lse.ac.uk/DARP

ABSTRACT

In this paper we propose to measure inequality of educational achievements by constructing a Gini index on educational attainments. We then use the proposed measure to analyse the relationship between inequality in incomes and educational achievements (in terms of both the average attainments and its concentration). Even if theoretical considerations suggest a non-linear relationship between these two measures of inequality, actual data indicate that average years of education have a stronger negative impact on measured income inequality. Multivariate regressions also prove that, once we take into account the negative correlation between average educational achievement and its dispersion, the relationship between income inequality and average years of schooling is U-shaped, with a lower turning point at 6.5 years. Income inequality is also negatively related to per capita income and positively related to capital/output ratio and government expenditure in education.

ACKNOWLEDGEMENT

The author thanks the comments of the participants at the workshop on “Inequality, growth and poverty under the Washington consensus” (Wider - Helsinki - July 1999) on an earlier version of this paper. He also thanks Meghnad Desai, David Soskice and an anonymous referee for helpful comments. Additional thanks to the audience of seminars held at Wissenschaftszentrum (Berlin), at the London School of Economics, at the SASE conference (London - June 2000), University of Pisa and University of Piacenza for helpful comments. Luca Flabbi provided excellent research assistantship. Financial support from the Ministry of University (MURST fondi ex-40%) is gratefully acknowledged.

Keywords: education inequality; income inequality; economic growth

JEL classification: 120; J24; 015; N30

Mailing address:
Daniele Checchi
Dipartimento di Economia Politica e Aziendale
Università degli Studi di Milano
via Conservatorio 7
20122 MILANO - Italy
tel. +39-02-76074-519(dir) 501(secr) fax +39-02-76009695
email daniele.checchi@unimi.it

EDUCATION INEQUALITY AND INCOME INEQUALITY †

Daniele Checchi - University of Milan - Italy

Mailing address:
Daniele Checchi
Dipartimento di Economia Politica e Aziendale
Università degli Studi di Milano
via Conservatorio 7
20122 MILANO - Italy
tel. +39-02-76074-519(dir) 501(secr) fax +39-02-76009695
email daniele.checchi@unimi.it

JEL classification: I20; J24; O15; N30
Keywords: Education inequality; Income inequality; Economic growth.

Abstract

In this paper we propose to measure inequality of educational achievements by constructing a Gini index on educational attainments. We then use the proposed measure to analyse the relationship between inequality in incomes and educational achievements (in terms of both the average attainments and its concentration). Even if theoretical considerations suggest a non-linear relationship between these two measures of inequality, actual data indicate that average years of education have a stronger negative impact on measured income inequality. Multivariate regressions also prove that, once we take into account the negative correlation between average educational achievement and its dispersion, the relationship between income inequality and average years of schooling is U-shaped,

† The author thanks the comments of the participants to the workshop on "Inequality, growth and poverty under the Washington consensus" (Wider - Helsinki - July 1999) on an earlier version of this paper. He also thanks Meghnad Desai, David Soskice and an anonymous referee for helpful comments. Additional thanks to the audience of seminars held at Wissenschaftszentrum (Berlin), at the London School of Economics, at the SASE conference (London - June 2000), University of Pisa and University of Piacenza for helpful comments. Luca Flabbi provided excellent research assistantship. Financial support from the Ministry of University (MURST fondi ex-40%) is gratefully acknowledged.

with a lower turning point at 6.5 years. Income inequality is also negatively related to per capita income and positively related to capital/output ratio and government expenditure in education.

Contents:

1. The issue

In the literature on the relationship between income inequality and output growth, several authors claim that greater income inequality reduces growth.[1] The empirical evidence indicates that one standard deviation decrease in income inequality raises the annual growth rate of product per capita by 0.5-0.8 percentage points. However, there is no consensus about the underlying causal mechanism. On one side, a political economy mechanism calls for a role for redistributive policies: greater income inequality generates increased social pressure and social instability, and this creates an adverse environment for investment in physical capital. On the other side, greater income inequality and greater poverty inhibit access to schooling and investment in human capital, thus reducing the potential for growth. Both explanations are at odds with a deeper scrutiny. The political mechanism hinges on the disincentive effect created by fiscal redistribution, which is not confirmed by the data.[2] The liquidity constraint explanation requires that the access to education be prevented by lack of financial resources, which is hardly the case in countries where public education is nearly cost-free at the compulsory level.[3] However both strands of literature make many simplifying

---

[1] Good surveys of this literature can be found in Benabou (1996a), Bourguignon (1996), Aghion, Caroli and Gracia-Peñalosa (1999), and Barro (1999).

[2] See Perotti (1996).

[3] Some empirical evidence in support of these propositions is offered in Bourguignon (1994), Checchi (1999) and Filmer and Pritchett (1998).

assumptions, the main one of which is that income inequality and educational choices are perfectly correlated and that the resulting earning distribution replicates educational choices. This allows the identification of an intergenerational equilibrium in income and education distributions. Since the two variables are perfectly correlated, the distribution of incomes and the distribution of human capital are shaped by the same factors. In many models, the same barriers (the absence of financial markets for education financing, the cultural poverty of the environment, the inefficiency of the public administration in tax levying) prevent investment in human capital by a fraction of the population, who subsequently earns less income.[4] Whenever there is some intergenerational persistence (via monetary inheritance or the effects of family cultural background), the very same portion of the population remains trapped at low levels of education and low levels of income for more than one generation. Thus, within the logic of formal models, illiterate people and the poor are synonymous. But in reality things are far more complicated. Educational choices are also correlated with the public provision of schools, the prohibition on children labour and the generally available opportunities in the labour market.[5]

[4] For example, Galor and Zeira (1993), Banerjee and Newman (1993) and Piketty (1997) consider financial market imperfection, while Benabou (1996a) takes into account the role of social capital, and Perotti (1993) points to the stage of development and the level of available resources.

[5] For example, in rural economies the output gains of child labour are the main obstacle to schooling among children. See the Zambian case described by Skyt Nielsen (1999), the

Analogously, income distribution can be more closely related to employment composition, labour legislation, trade union coverage, and fiscal policies than to educational achievements among the population.[6]

However, the distribution of incomes and the distribution of educational attainments are obviously related. On one hand, income inequality may prevent access to education when education is too costly for the family: the more skewed the income distribution, the higher the population share excluded from schooling and the higher the inequality in educational achievements. In this respect, we have a self-perpetuating poverty trap that can only be avoided by easing access to education.[7] On the other hand, improved access to education raises the earning opportunity of the lowest strata and, other things being constant, reduces earnings inequality. As long as total income is proportional to labour income, we can expect a positive correlation between the distribution of educational achievements and the distribution of incomes in the population. But the 'other things being constant' assumption is rather

---

Bangladesh case analysed by Ravallion and Wodon (1999) and the Indian case discussed by Weiner (1991).

[6] See Gottschalk and Smeeding (1997) and Bardone, Gittleman and Keese (1998) for the determinants of earnings distribution in OECD economies. Globalization and the effect on wage inequality are discussed in Borjas and Ramey (1995), Sachs and Shatz (1996) and Feenstra and Hanson (1996).

[7] Checchi (1999) shows that income inequality effectively reduces school enrolment, mainly at secondary level. Similar results are in Flug, Spilimbergo and Wachtenheim (1998). From a formal point of view, this corresponds to the case where *current* income inequality affects the *rate of change* of inequality in educational achievement.

crucial here, since we have to take into account the general equilibrium consequences of these changes. Consider for example the case of skill-biased technological change. Many authors agree that this is one of the potential reasons for the boost in the college ‘premium’, at least in the United States. With a time lag, this has produced an increase in college enrolments despite the rise in tuitions. Until the supply of new college graduates depresses the premium, we will observe growing income inequality, accompanied by a reduction in inequality in educational achievement.[8]

In the present paper we are concerned with the existing correlation between the distribution of educational achievements and the distribution of incomes. The literature we have mentioned so far does not provide a clear-cut prediction on the sign of this relationship. For this reason, in this paper we intend to investigate the empirical determinants of aggregate income inequality and, more specifically, the relative contribution of education to measured income inequality. In our opinion, this is crucial for two considerations. First, from a theoretical point of view, it is important to understand the plausibility of studying intergenerational equilibria under stationary distributions of income and human capital in the

[8] Freeman (1986) has shown the existence of a similar phenomenon during the 1960s for engineers in the US and has provided a ‘cob-web’ model for the dynamics of this phenomenon. For more recent evidence, see Murphy, Riddel and Romer (1998). Galor and Tsiddon (1996) present a model where the opening of access to education initially widens and then reduces the skill premium.

population. Second, and far more important, from a policy point of view we want to understand whether urging countries (or people) to increase their educational achievements is going to exacerbate, moderate, or have little influence on the subsequent earnings distribution.

The paper is organized as follows. The next section reviews the literature on income inequality determinants. The third section provides empirical evidence. The fourth section concludes. Appendix I indicates data sources and discusses data reliability. Appendix II proves some proposition reported in the text.

2. Existing literature

There is a growing literature on the current trends in income inequality at world level.[9] Rising income inequality occurred initially in Anglo-Saxon countries, but now is affecting most industrialized nations.[10] Among the potential causes of this phenomenon, the reduction of the redistributive role of the state, the decline in union presence in the workplace, the increased competition at international level, technological progress and all possible combinations of these are often indicated. However, the experiences at national level are very

---

[9] See Atkinson (1999), Cornia (1999) and the references therein.

[10] Milanovic (1999) has computed an increase of 3 Gini points in world income inequality from 1988 and 1993, mainly attributable to between-country inequality.

diversified, and it is quite hazardous to draw general conclusions. Apart from the Kuznets (1955) hypothesis on the existence of a non-linear relationship between output per capita and income inequality, we do not find much progress in the statistical explanation of the observed inequality. In particular, little work has been undertaken so far seeking to test alternative explanations of the evidence on income distribution and even less concerning the relationship between educational attainment and income inequality. This is surprising, given the fact that compulsory education is publicly and freely provided in almost all countries of the world.

At the best of our knowledge, the existing literature on the effects of educational attainments on income inequality mainly focuses on the two first moments of education distribution, namely the average educational attainment (average years of schooling) and the dispersion of schooling in the population. For the first, Barro (1999) suggests that the relationship between income inequality and output growth is negative for poor countries and positive for rich countries, the threshold being a gross domestic product per capita lower than $2.070 at 1985 prices.[11] He runs conditional convergence regressions on the income inequality (from the Deininger and Squire, 1996, dataset) measured five years earlier in order to exclude the case

[11] Perotti produced some evidence pointing in the same direction as discussant of Benabou (1996b).

of reverse causation. Then he moves this regressor to the left-hand side and studies the determinants of income inequality. He puts forward some evidence on the existence of an inverted-U-shaped relationship between output per capita and income inequality (with a turning point around $1.636). He controls for educational achievement by introducing average educational attainments at three levels (primary, secondary and tertiary).[12] But his results are difficult to interpret in this respect, because of the contemporaneous presence of different information on the distribution of educational achievements (namely, the contribution of average human capital and its distribution across population subgroups).[13]

A similar strategy is followed by O'Neil (1995), who decomposes output growth over 1967-85 into a 'quantity' component (as measured by enrolment rates) and a 'price' component (as measured by relative stocks of human capital). His analysis suggests that, while there is convergence among countries in the level of educational achievement, the price effect works in the opposite direction.[14] In the same line of

[12] 'The panel also includes the average years of school attainment for people older than 15, classified over three educational levels, primary, secondary and higher. The results are that primary schooling is negatively and significantly related to inequality, secondary school is negatively (but not significantly) related to inequality, and higher education is positively and significantly related to inequality' (Barro, 1999: 26).

[13] To be more precise: an additional average year in either primary school, or in college should raise average educational achievement, but the former reduces educational variance in the population, while the latter raises it.

[14] 'The results in table 2 also show that, for both developed countries and Europe, the rise in the return to education experienced over the last two decades has caused incomes to

research, Deininger and Squire (1998) show that initial inequality in assets (land) is relevant in predicting both income growth and changes in income inequality.[15] Since land inequality also reduces average years of education in their regressions, they explain this evidence by referring to the liquidity constraints on access to education. As a consequence, income inequality and educational attainments are positively correlated because of the presence of a third conditioning variable (wealth inequality). However, while asset (or income) inequality may reduce the creation of new human capital (the 'flow' represented by new school-leavers), we see no good reason to suppose it might depreciate existing human capital (the 'stock' represented by the average educational attainment of the population).[16] In a related paper, Li, Squire and Zou (1998) interpret the evidence that the effect of (initial-period) average secondary school years on income inequality is significant as a proxy for a political effect: the more political freedom there is, the more informed is society, the more difficult it will be for the rich to appropriate extra resources.[17]

---

diverge substantially, as those countries that are better endowed with skilled labor reap the benefit of the rising premium' (O'Neil 1995: 1,295).

[15] 'Low initial inequality is thus doubly beneficial. It is associated with higher aggregate growth, the benefits of which accrue disproportionately to the poor' (Deininger and Squire 1998: 261).

[16] In addition, their analysis involves only 52 observations, and liquidity constraints are represented mainly not by land distribution, but by the level of current income.

[17] Gradstein and Milanovic (2000) provide additional evidence on the potential existence of links between political inclusion and income equality. However, it is not clear which is the direction of causation: whether extended franchise supports more redistributive policies, or whether less unequal societies strengthen democracy. Justman and Gradstein (1999) present similar ideas through a formal model that predicts the existence of an

All these papers recognize the existence of a distributional aspect in the relationship between income inequality and educational inequality, but they rely mainly on average attainments. In contrast, the issue of education distribution is central in the paper by Lopez, Thomas and Wang (1998).[18] They demonstrate that human capital, as measured by average educational attainment, is statistically non-significant in output-growth regressions unless one does not control for the distribution human capital ('who gets what') or for openness to international trade ('what to do with education'). They explain their evidence (on 12 countries over 1970-94) through reference to the absence of tradability in human capital that makes price equalization impossible and can produce shortages in human capital during physical capital accumulation. Along the same lines is the argument by Higgins and Williamson (1999), who predict the Gini index of income inequality using output per worker (linear and quadratic, in accordance with the

---

inverted-U-shaped relationship between income inequality and franchise. When the median voter income exceeds the average income, regressive redistribution policies are adopted, and inequality rises; as long as the median voter income remains below the average income, progressive redistributive policies tend to be adopted. Finally, Breen and García-Peñalosa (1999) find that greater income inequality is positively associated with higher income volatility, as measured by standard deviation in output growth rates, and they show that this finding is robust even if one controls for previous variables. They suggest that this could be due to the fact that firms offer an implicit contract to risk-averse workers. When the environment becomes more uncertain, the cost of this implicit insurance rises, and wages are consequently reduced, thus increasing income inequality.

[18] More recently, Castello and Doménech (2001) obtain analogous empirical results measuring educational inequality by means of Gini concentration index. Galor and Tsiddon (1997) offer another theoretical paper focusing on educational inequality as a source of technological progress (and output growth).

hypothesis of Kuznets) and cohort-size effects (large mature working-age cohorts are associated with lower aggregate inequality because of relative excess supply). However, as they explicitly recognize, this approach neglects the endogeneity of educational choices. Let us suppose that a society is undergoing a transitional phase, in which the average educational requirement is rising, such that the younger cohorts are better educated than the older ones. Other things being constant, the smaller the size of the more well educated cohort, the lower the recorded inequality in incomes. It is therefore rather possible that, through reliance on age-composition variables, the authors were actually capturing educational changes.[19] At any rate, the two measures for educational achievement (average educational attainment and some measure of the dispersion of attainment) are intertwined. Both Ram (1990) and Londoño (1996) claim the existence of an inverted-U-shaped relationship between educational achievement and educational inequality, and they locate the turning point at 6.8 average years of education.[20] However, they do not provide a sound theoretical

[19] It is true that they control for secondary enrolment rates, but, as we have already argued above, this variable measures the flow and not the distribution of the stock of human capital.

[20] 'In a society where there is no education for everyone, the level of education is zero and the variance of education among the population is naturally zero. In a society where the entire population reaches the maximum level of education, the level of education is at maximum, but the variance, again, is zero. . . . In the interim period, the variance of education tends to rise with the increase in the level of education until it reaches a turning point, after which it decreases' (Londoño, 1996: 13). However, this reasoning is not rigorous on statistical grounds since a generalized increase of education in the population produces an increase in average achievement without necessarily raising educational inequality.

argument to explain this occurrence, nor do they show whether this relationship might hold for alternative measures of dispersion or concentration.[21]

What do we learn from this literature? Three points seem unquestionable, at least in the empirical literature:

i) income inequality is clearly related to the stage of development in accordance with some sort of Kuznets relationship;

ii) income inequality also reflects the skill level of the population, as proxied by average educational attainment.

iii) it is still unclear whether and how average educational attainment and dispersion of schooling jointly contribute to shape income distribution.

In addition, in all previous work, we have not found any measure related to labour market institutions (such as the presence of unions, unemployment benefits, or the minimum wage).[22] In the sequel, we intend to shed some light on the third point by jointly considering average educational achievement and dispersion in the population in predicting income inequality. We will also take into account some measures of the quality / quantity of the resources invested in education.

---

[21] See Appendix II where this issue is extensively discussed.

[22] Nor do we find controls for inequality in explaining employment / unemployment rates. See Glyn and Salverda (2000) and Kahn (2000) for an analysis of OECD countries in this respect.

3. Available measures of inequality

Starting from enrolment rates and making appropriate assumptions about mortality rates, Barro and Lee (1996) provide estimates of the human capital stock of a country. Using mild assumptions on the demographics (similar to the permanent inventory method used to estimate the stock of physical capital), starting from enrolment rates and possessing the distribution of educational achievement at some reference time-point, one can obtain estimates of the average years of education among the population for each level of education. Let us illustrate this with an example. Consider a population in which each age cohort grows at a constant rate n and in which the probability of death is constant across ages and equal to δ. If we define k as the life expectancy,[23] and denote with $Pop_{t,j}$ the population aged j at time t, the entire population is given by

$$Pop_t = Pop_{t,k} + Pop_{t,k-1} + ... + Pop_{t,0} =$$
$$= Pop_{t-k,0}(1-\delta)^k + Pop_{t-k,0}(1-\delta)^{k-1}(1+n) + ... + Pop_{t-k,0}(1+n)^k = Pop_{t-k,0}\sum_{i=0}^{k}(1-\delta)^{k-i}(1+n)^i$$

Suppose that schooling consists of one year and dropout rates are zero (such that enrolment rates coincide with graduation rates). Under this assumption, if we indicate with π_t the

[23] It can be determined as $k:(1-\delta)^k \approx 0$.

percentage of population born in t that achieves education, we obtain the number of people with education as

$$Pop_t^{educated} = \pi_{t-k} Pop_{t,k} + \pi_{t-k+1} Pop_{t,k-1} + ... + \pi_t Pop_{t,0} = Pop_{t-k,0} \sum_{i=0}^{k} \pi_{t-i} (1-\delta)^{k-i} (1+n)^i$$

Therefore, under previous assumptions the current population share with education is given by

$$HC_t = \frac{Pop_t^{educated}}{Pop_t} = \frac{\pi_{t-k} Pop_{t,k} + \pi_{t-k+1} Pop_{t,k-1} + ... + \pi_t Pop_{t,0}}{Pop_{t,k} + Pop_{t,k-1} + ... + Pop_{t,0}} = \frac{\sum_{i=0}^{k} \pi_{t-k+i} (1-\delta)^{k-i} (1+n)^i}{\sum_{i=0}^{k} (1-\delta)^{k-i} (1+n)^i} =$$

$$= \frac{\sum_{i=0}^{k} \pi_{t-k+i} \left(\frac{1-\delta}{1+n}\right)^{k-i}}{\sum_{i=0}^{k} \left(\frac{1-\delta}{1+n}\right)^{k-i}} = \sum_{i=0}^{k} \pi_{t-k+i} \frac{\omega^{k-i}}{\sum_{j=0}^{k} \omega^{k-j}} = \sum_{i=0}^{k} \pi_{t-k+i} \omega^{k-i} \left[\frac{1-\omega}{1-\omega^{k+1}}\right], \qquad \omega = \left(\frac{1-\delta}{1+n}\right) < 1 \tag{1}$$

which is a weighed average of past enrolment rates (with declining weights, as in an Almon's polynomial) . In the particular case of constant enrolment rates (i.e. $\pi_i = \pi, \ \forall i$), equation (1) collapses to $HC = \pi$.[24] Repeated application of equation (1) yields

[24] With educational cycles lasting more than one year and positive dropout rates, things are more complicated, but the logic of the argument holds unchanged. Indicating with λ_t the age cohort share enrolling a school level lasting n years (say primary school starting at the age of m and lasting n years), with a (constant) dropout rate of γ, then the enrolment rate would be

$$\pi_t = \frac{\lambda_{t-n} (1-\delta)^{m+n} (1+n)^{k-m-n} (1-\gamma)^n + \lambda_{t-n+1} (1-\delta)^{m+n-1} (1+n)^{k-m-n+1} (1-\gamma)^{n-1} + ... + \lambda_t (1-\delta)^m (1+n)^{k-m}}{(1-\delta)^{m+n} (1+n)^{k-m-n} (1-\gamma)^n + (1-\delta)^{m+n-1} (1+n)^{k-m-n+1} (1-\gamma)^{n-1} + ... + (1-\delta)^m (1+n)^{k-m}}$$

$$HC_t = (\omega HC_{t-1} + \pi_t)\left[\frac{1-\omega}{1-\omega^{k+1}}\right] = (\omega HC_{t-1} + \pi_t)\Omega, \quad \Omega < 1 \qquad (2)$$

If we now indicate with HC_{pt} the population share with some primary education and P_{pt} as the enrolment rate for primary education, both measured at time t, it is easy to understand why the former variable can be thought as the integral of the latter (using the decline rate $\mu = 1-\omega$ as a discount factor). In symbols

$$\begin{aligned} HC_{pt} &= \Omega\{[HC_{p0} \cdot (1-\mu) + P_{p1}] \cdot (1-\mu) + P_{p2}\} \cdot (1-\mu) + \ldots\ldots = \\ &= \Omega\left[HC_{p0} \cdot (1-\mu)^t + \sum_{i=1}^{t} P_{pi} \cdot (1-\mu)^{t-i}\right] \end{aligned} \qquad (3)$$

where HC_{p0} is the (estimated) population share with primary education at a given year of reference (usually a census year), and μ represents the (constant) decline rate of an age cohort in the population. The use of a continuous time representation yields:

$$HC_{pt} = \Omega\left[HC_{p0} \cdot (1-\mu)^t + \int_0^t P_p(s) \cdot \exp(-\mu \cdot s)ds\right] \qquad (4)$$

Should the growth rate of the population or the mortality rate not remain constant over the years, the above derivations do not correspond exactly to the theoretical value implied by equation (4). By multiplying HC_p by the number of years

---

which in turn is a weighed average of enrolment at first year, taking into account the required to complete primary education, we obtain the average number of years of primary education in the population. When we possess this piece of information for each level of education, we have an approximation of the distribution of the human capital stock in a country. The calculation is just an approximation because in many cases an attained educational level, say, a secondary degree, may actually be acquired after a longer-than-average period of study (because of repetition); in addition we could encounter people who attended schools without attaining any certificate because earlier dropping-out. Even if the information on dropout rates is available, we may not know when individuals leave a course of study; therefore, we cannot integrate this information in the computation of the average stock of human capital.[25] Once we have the rough distribution of educational achievement in the population, it is possible for us to calculate several measures of inequality, among which the Gini concentration index of the distribution of attained education is one of the easiest to compute. If only subgroup averages are known, the general definition of the index is modified accordingly:

$$G = \frac{1}{2n^2 \cdot \mu} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| n_i - n_j \right| = \frac{1}{2\mu} \sum_{k=1}^{M} \sum_{h=1}^{M} \left| \bar{n}_k - \bar{n}_h \right| \cdot HC_k \cdot HC_h \qquad (5)$$

---

decline due to dropout.

[25] Dropout rates are effectively available in Barro and Lee data set at the primary level. This variable ranges from an average (over the period 1960-95) of 3.35% in OECD countries to 39.8% in Sub-Saharan Africa, 39.7% in South Asia and 36.6% in Latin America.

where N is the population size, n_i is the number of years of schooling of individual i, μ is the average years of schooling in the population, M is the number of subgroups and $\bar{n}_h$ is the (average) educational attainment in subgroup h. In the case of educational attainments, Barro and Lee (1996) provide us with the available information on three educational levels.[26] This allows us to divide the population in four subgroups: higher education (a share HC_h has attained n_h years of education), secondary education (a share HC_s with n_s years), primary education (a share HC_p with n_p years) and a residual group without education ($HC_n = 1 - HC_h - HC_s - HC_p$, for which zero education is assumed).[27] By construction, the average population attainment is given by

$$\mu = \overline{HC} = HC_p \cdot n_p + HC_s \cdot n_s + HC_h \cdot n_h \tag{6}$$

and the Gini index on educational attainments is computed as follows

[26] This is another obvious approximation, since we are standardizing educational systems into a tripartite classification, corresponding to UNESCO 'ISCED' (international standard classification system of education levels) standards. However, if a country (like Germany) has double-track secondary education (high school and vocational training), each with a different duration, the duration will nevertheless be computed as a though it were a single figure.

[27] Barro and Lee (1996) make a distinction between 'attained' and 'completed' educational levels. Given the high correlation between the two series, we have preferred to adopt the former variable because there are fewer missing observations for it.

$$G_{\mathrm{ed}} = \frac{HC_h \cdot n_h \cdot \left(HC_s + HC_p + HC_n\right) + HC_s \cdot n_s \cdot \left(-HC_h + HC_p + HC_n\right) + HC_p \cdot n_p \cdot \left(-HC_h - HC_s + HC_n\right)}{\overline{HC}} =$$

$$= HC_n + \frac{HC_h HC_s \left(n_h - n_s\right) + HC_h HC_p \left(n_h - n_p\right) + HC_s HC_p \left(n_s - n_p\right)}{\overline{HC}} \qquad (7)$$

Starting from the original Barro and Lee (1996) dataset, we have extended the observations up to 1995. We therefore have information about educational achievements in the population for 149 countries at five-year intervals over the period 1960-95. Overall, these data cover three-fourths of the 210 countries listed by the World Bank (1998), but account for 86.3 per cent of the world population (in 1990). However, missing values have reduced the potential number of observations from 1,192 to 848 cases, corresponding to 117 countries (with an average of 7.2 observations per country). Descriptive statistics on these variables appear in table 1 at world aggregate level, in table 2 with temporal disaggregation and in table 3 with temporal and regional disaggregation; additional information about data sources is in Appendix 1.

In the most recent year of observation (1995), we find that one-third of the world population is illiterate; one-third has primary education, and the remaining one-third has secondary schooling or more. During the time span under consideration in this paper, the average number of years of education rose from 4.3 to 5.8 at the world level, although this rise was accompanied by growing gaps in the same variable computed at regional level. The population share composed of illiterate people or people with primary education exhibited a declining trend, with some reversal at the end of the period, and there was a similar trend in the index of inequality of educational achievement. But the global picture varied by region: while educational inequality declined in North Africa, South Asia and the formerly planned economies, it decreased during the first three decades, but rose thereafter in other regions (especially in sub-Saharan Africa). Inequality in terms of years of schooling remained almost constant at low levels in the OECD countries, despite the increase in the average educational attainment. It is therefore difficult to trace out a single trend at world level, especially because there seems to be a difference among countries in the rates of change, as well as in the levels of the variables.

Table 1 – Descriptive statistics

Variable	*Variable name*	*Mean (weight=pop)*	*Median (weight=pop)*	*Std.Dev. (weight=pop)*	*Obs.*
Population share without education	HC_n	40.4%	43.1%	0.278	883
Population share with primary education	HC_p	33.8%	32.3%	0.172	902
Population share with secondary education	HC_s	19.8%	17.2%	0.143	916
Population share with higher education	HC_h	5.6%	2.5%	0.077	919
Average duration of primary education	n_p	5.35	5.10	1.153	869
Average duration of secondary education	$n_s - n_p$	4.59	4.58	0.824	929
Average duration of higher education	$n_h - n_s$	3.49	3.33	0.791	898
Average years of education	μ	4.66	3.89	2.757	848
Gini index on educational attainment inequality	*Ginied*	49.32	51.74	23.261	848
Gini index on income inequality	*Gini*	38.01	36.85	8.239	546

Table 2 – Mean values (weight=population) across years

Variable	*1960*	*1965*	*1970*	*1975*	*1980*	*1985*	*1990*	*1995*
Population share without education	46.3%	46.7%	44.1%	44.6%	43.1%	38.6%	33.5%	35.5%
Population share with primary education	38.1%	37.1%	37.4%	34.8%	31.2%	32.6%	33.2%	32.3%
Population share with secondary education	12.5%	12.7%	14.0%	16.2%	20.4%	22.5%	25.3%	22.9%
Population share with higher education	2.5%	2.8%	3.5%	4.4%	5.4%	6.3%	8.1%	9.0%
Average duration of primary education	4.91	5.02	5.11	5.05	5.22	5.34	5.40	6.37
Average duration of secondary education	4.45	4.53	4.65	4.61	4.47	4.52	4.59	4.85
Average duration of higher education	3.21	3.75	3.45	3.55	3.45	3.40	3.41	3.70
Average years of education	4.31	3.67	3.93	3.92	4.30	4.81	5.39	5.86
Gini index on educational attainment inequality	44.89	53.63	52.43	53.71	52.07	48.38	44.31	47.03
Gini index on income inequality	42.05	36.65	37.14	36.47	37.65	37.67	38.43	39.35

Table 3 – Mean values (weight=population) across years – regional variations

Variable	*1960*	*1965*	*1970*	*1975*	*1980*	*1985*	*1990*	*1995*
OECD countries								
Average years of education	6.75	6.98	7.46	7.65	8.59	8.66	9.00	8.81
Gini index on educational attainment inequality	20.68	21.41	21.26	22.64	20.75	20.72	20.98	24.21
Gini index on income inequality	39.55	37.27	38.01	36.87	35.87	36.20	36.35	37.36
North Africa and the Middle East								
Average years of education	1.03	1.12	1.36	1.57	2.14	2.77	3.48	4.90
Gini index on educational attainment inequality	85.95	86.03	83.38	83.21	77.70	71.00	64.82	52.71
Gini index on income inequality	49.05	46.87	49.59	49.29	41.37	47.40	38.72	35.30
Sub-Saharan Africa								
Average years of education	1.01	1.65	1.61	1.66	1.96	2.14	2.32	2.74
Gini index on educational attainment inequality	82.47	74.39	74.83	72.79	67.08	64.33	63.08	75.35
Gini index on income inequality	51.86	50.76	56.22	44.31	42.47	46.24	52.75	44.98
South Asia								
Average years of education	0.91	1.37	1.74	2.08	2.45	2.81	3.20	4.23
Gini index on educational attainment inequality	86.23	79.67	77.99	76.14	76.71	72.78	69.08	61.49
Gini index on income inequality	38.90	37.40	36.74	38.37	38.22	38.64	35.52	30.02
East Asia and the Pacific								
Average years of education	3.72	3.96	4.34	4.71	5.35	5.82	6.31	6.43
Gini index on educational attainment inequality	50.64	49.02	41.24	39.11	35.33	31.86	31.44	39.27
Gini index on income inequality	40.19	37.51	36.41	39.65	39.18	39.88	40.02	38.38
Latin America and the Caribbean								
Average years of education	3.06	2.99	3.37	3.47	3.97	4.13	4.74	6.17
Gini index on educational attainment inequality	49.70	50.75	47.68	45.05	44.27	44.23	39.08	43.22
Gini index on income inequality	52.22	49.93	53.99	53.77	52.31	54.66	54.63	56.05
formerly Centrally Planned Economies								
Average years of education	3.92	4.83	5.28	3.61	3.68	4.96	6.09	8.17
Gini index on educational attainment inequality	33.37	35.72	32.20	56.04	52.86	44.69	35.15	23.12
Gini index on income inequality	===	30.52	27.83	26.72	32.06	30.50	33.37	41.53

Since we are interested in the relationship between educational achievement and income distribution, we now add the dynamics of income inequality to the picture. Here, we rely on the dataset of Deininger and Squire (1996) and on the 'World Income Inequality Dataset' (WIID) collected by WIDER, both of which contain a substantial amount of information on inequality measures collected from secondary sources. Among these measures, the Gini index on income inequality is the most readily available.[28] In the present case, we have information on 546 observations, corresponding to 113 countries (with an average of 4.8 observations per country). If we restrict our selection to the subset in which there is information about both income and education inequality, we have 478 observations for 97 countries (with an average of 4.9 observations per country; Appendix 1 contains a list of the countries). Table 2 reports the population-weighted average for this measure computed on all available information in the dataset.[29] We notice that, despite a declining trend in educational inequality (reversed only during the 1990s), income inequality at world level started rising after 1975. Figure 1 (which graphs the data reported in Table 3), seems to indicate that this is mainly attributable to the OECD, the Latin American countries and the formerly centrally planned economies.

One may wonder whether missing values may distort the sample. Looking at Table 4 we may notice that effectively we observe an over representation of richest countries and an under representation of the poorest countries, both mainly attributable to income inequality data. If we run probit

[28] See Appendix 1 for a discussion of the changes made in the original Deininger and Squire dataset, including the 1995 update of the observations.

[29] Given that the Gini index is not decomposable by population subgroup, the trend in the population-weighted average has to be viewed with caution. See Milanovic (1999) for a more accurate picture based on population surveys (albeit with observations only over two years, 1988 and 1993).

regression on data availability (as it is done in Table A1 in the Appendix) we find that under-representation involves the last observation, the less populous and the poorest countries, especially in the North-African region. On the contrary, Latin American and Asian countries tend to be overepresented.

Table 4 – Available observations for world regions

World regions	potential sample	missing income	missing education	effective sample	missing income	missing education	effective sample
OECD countries	192	51	17	134	26.6%	8.9%	69.8%
North Africa and the Middle East	144	110	44	25	76.4%	30.6%	17.4%
Sub-Saharan Africa	352	256	166	58	72.7%	47.2%	16.5%
South Asia	56	24	3	31	42.9%	5.4%	55.4%
East Asia and the Pacific	120	53	39	65	44.2%	32.5%	54.2%
Latin America and the Caribbean	232	112	48	118	48.3%	20.7%	50.9%
formerly Centrally Planned Economies	96	40	27	47	41.7%	28.1%	49.0%
Total	1192	646	344	478	54.2%	28.9%	40.1%

4. Empirical analysis

We now move to the theoretical investigation of correlation between the distribution of education and the distribution of incomes. So that we can make more precise statements, let us now consider what we may expect from theoretical models. If we adopt a standard version of the theory of human capital investment, initially proposed by Becker (1964) and subsequently taken up by Mincer (1974) to estimate the returns to education, the (log)incomes and years of education are linearly related. In fact, when a Mincer-Becker theory of earnings applies, individual earnings would be determined as:

$$\log(y_i) = \alpha + \beta \cdot n_i +$$

$$+ \text{individual characteristics (gender, age, experience, etc)} + \varepsilon_i \qquad (8)$$

where y_i is the earning capacity of individual *i*, n_i is the educational attainment of individual *i* (measured in years of schooling), β is the (percentage) rate of return to education and α is the earning of an individual without formal education; ε_i is an error term assumed to be i.i.d. (identically independently distributed). If we assume that the *individual characteristics* are idiosyncratic in the population and orthogonal with acquired education, population subgroups differ only in terms of average educational achievement (namely, the within-group variance is constant).If we assume that the *individual characteristics* are idiosyncratic in the population and orthogonal with acquired education, population subgroups differ only with average educational achievements (namely the variance within group is constant)[30]. We therefore expect there to be a relationship between the distribution of educational achievements and the distribution of actual incomes. However, the things are not so simple. Inserting equation (8) into equation (7) and ignoring the (average) individual characteristics, we obtain the Gini index of log-income inequality as:

[30] Actually, Mincer (1996) shows that between-group variance in earnings distribution in the US remained nearly constant during 1970-90, whereas the within-group variance expanded after the 1980s.

$$G_{\log-\text{income}} = \frac{HC_h \cdot \beta \cdot n_h \cdot \left(HC_s + HC_p + HC_n\right) + HC_s \cdot \beta \cdot n_s \cdot \left(-HC_h + HC_p + HC_n\right) +}{\alpha + \beta \cdot \overline{HC}} \tag{9}$$
$$\frac{+ HC_p \cdot \beta \cdot n_p \cdot \left(-HC_h - HC_s + HC_n\right)}{\alpha + \beta \cdot \overline{HC}}$$

or more synthetically

$$G_{\log-\text{income}} = \frac{\overline{HC}}{\frac{\alpha}{\beta} + \overline{HC}} \cdot G_{ed}\left(\overline{HC}\right) = f\left(\underset{\pm}{\overline{HC}}\right) \tag{10}$$

Equation (10) suggests that, at a given average in educational achievements, the inequality in education and the inequality in (log)earnings are linearly related. If incomes are proportional to earnings, this also applies to inequality in (log)incomes. However, since the inequality in education is negatively related to average education, the actual relationship is non-linear.[31] The situation is rendered more complicated by the fact that we do not possess individual data allowing the calculation of inequality measures for (log)incomes. Rather, we are forced to rely on aggregate measures based on actual incomes. Once more, the relationship between the inequality measures obtained from actual values of the variables and the corresponding measures computed based on the logarithms is

[31] In a previous version of this paper, we made use of simulations (relying on the observed values for educational achievement) to analyse the relationship between education and income inequality under the assumption that returns to education are constant. We found that the relationship is positive and stronger in countries with low-to-middle inequality in education (lower than 45%), whereas the same relationship is negative in countries with very high inequality in education. This is because the Gini concentration index is scale invariant (that is, it does not vary when we change the unit of measure), but not translation invariant. Therefore, given the presence of a constant ($\alpha \neq 0$), a generalized rise in educational achievement (at the given inequality in educational attainments) induces a change in income inequality.

not easily ascertained (unless we impose more structure to the problem by assuming a specific functional form for the frequency distribution). However, it can be formally demonstrated that – under mild assumptions about the distribution of education in the population and the general assumption that the rate of return to education is constant – the relationship between the Gini index of actual incomes and the average years of education initially rises and then declines.[32] When the assumptions hold, income inequality, education inequality and average educational inequality are strictly related, as shown in Figure 2, where we have also added a fourth variable, the output per capita, in order to control for an exogenous driving force.

Starting from the lower right-hand quadrant, we assume that an increase in per capita income is associated with an increase in the average educational attainment. By construction, this yields a consequent decline in educational inequality (lower left quadrant).[33] If the relationship between average educational attainment and income inequality is non-linear, this necessarily implies a non-linear relationship between income inequality and education inequality (upper left quadrant). By the same token, we also obtain an inverted-U-shaped relationship between income inequality and per capita income, in the

[32] See the Appendix 3.

[33] See, again, the Appendix 3.

Kuznets tradition (upper right quadrant). The graph tells us a story about the transition from an uneducated population to an actual level of schooling. When only the élite attends schools, the average level of human capital development among the population is low, whereas the inequality in educational achievements and in incomes is high. A lowering of the access barriers to education leads to an initial increase and then to a decline in both inequality measures, and this is accompanied by a rise in average educational attainments. A similar story is obtained in the model presented by Galor and Tsiddon (1996).

A first inspection of our dataset indicates that this story may have some plausibility. Figure 3 gathers together all the available observations, whereby income inequality is measured by regression residuals on regional dummies and year-related dummies in order to compensate for trends in the variables and regional disparities. In addition to a mildly non-linear relationship between inequality in actual incomes and inequality in education (see the upper left quadrant), a similar relationship emerges between the former variable and (the log of) GDP per capita, in line with the Kuznets tradition (upper right quadrant). Without concerning ourselves too much about the direction of the causal relationship, we also find evidence of a strict positive correlation between output per capita and educational achievement (lower right quadrant). Finally, by construction, we find an inversely proportional relationship between inequality in education and average educational achievement (lower left quadrant).[34]

However, the dispersion of single observations suggests that many other forces are at work. We should not forget that the validity of the story of Figure 2 is conditional on the assumption that individual incomes are determined according to Becker's theory of human capital investment and that returns to education are constant and are, moreover, identical throughout the population and across countries. In reality, we know that earnings distribution is shaped by many other factors, including technology, unemployment rates, minimum wages, age composition, the existence of labour unions, and so on.[35] Were it certain that these factors remained constant during our sample period, we could consider them country-specific fixed effects. The problem is that there is no guarantee that they remained constant, especially if we take into account the

---

[34] However, the way we measure educational inequality is crucial. Had we chosen the standard deviation of educational achievement like Ram (1990), Londoño (1996) and IDB (1998), the relationship between average educational attainment and educational inequality would have appeared non-linear:

$$St.Dev._{ed} = \underset{(24.4)}{1.72} + \underset{(19.6)}{0.644} \cdot \overline{HC} - \underset{(17.8)}{0.056} \cdot \overline{HC}^2, \quad R^2 = 0.32, \qquad n = 848$$

In such a case, the turning point would occur at 5.75 years (rather than the 6.8 years measured by Ram, 1990).

[35] See Neal and Rosen (1998) for a general presentation of determinants of earnings distribution. Higgins and Williamson (1999) find evidence of an effect of age composition (as measured by the share of individuals aged 40-59 in the labour force) in determining income inequality. With reference to OECD countries, Bardone, Gittleman and Keese (1998) show that labour market institutions changed during the sample period: trade union density and coverage declined (especially within the Anglo-Saxon world), while minimum wages declined almost everywhere, and cuts to welfare assistance may have induced lower reservation wages.

transformation in public policies induced by the ‘transatlantic consensus’ (Atkinson, 1999).

As a consequence, instead of pretending to predict the shape and the evolution of income distribution worldwide, we follow in the sequel the less ambitious aim of discovering whether the average educational achievement and the distribution of educational attainment have played any role in determining income inequality. We have already mentioned the fact that other authors (Londoño, 1996; Deininger and Squire; 1998; Barro, 1999) have shown that average educational achievement is one of the determinants of actual income inequality. To this result, we now add an examination of the effect on income inequality of the distribution of educational achievement in the population.

In order to take into account the simultaneous effects of all the variables, we resort to multivariate regressions. We take our dataset as an unbalanced panel with a potential dimension of 752 observations (94 countries times 8 observations per country), which we reduce to 454 observations because of missing data on one or the other variable. Table 5 shows estimates of actual income inequality using fixed effects, whereas Table A2 in Appendix 2 relies on random effect estimators. In both tables we start with two alternative specifications of the relationship between income inequality and output per capita, without taking into account educational factors (first and second columns). Both specifications reject the hypothesis of a non-linear relationship between income inequality and per capita output.[36] The two measures are negatively correlated, with a rather low elasticity (-0.049 at sample means). This implies that, in order for the Gini index of income inequality to be reduced by 1 point, income per capita has to rise by $2,311 (at 1985 international prices). If we replace per capita income by educational variables (third and fourth columns), we notice an increase in explanatory power only if we consider average educational achievement. This is not surprising given the high correlation of the latter measure with per capita income. Both average educational achievement and educational inequality are significant, but the relationship between the two measures of inequality is opposed to the theoretical expectation (being U shaped and not inverted-U shaped). We consider GDP per capita and educational variables together in the fifth column. Here, we find that output per capita has a low negative impact, as does average human capital, though with a higher effect: an average increase of one year of education in the population lowers the Gini index of income inequality by more than 1 point. The sixth column offers an alternative (hyperbolic) specification of the functional relationship relating income inequality and average human capital: given the non-linear relationship existing between the Gini index of educational inequality, the variable $1/h_c$ seems

[36] On a different data set Anand and Kanbur 1993 obtained statistically significant results.

able to capture all the explanatory power contained in the educational distribution variable.[37]

However, the explanatory contribution of the distribution of educational achievement is region specific and rather unstable. If we run fixed effect estimates for world regions (as in Table 6), we see that education distribution variables are significant for OECD countries, Latin American and Sub-Saharan countries, that is the countries at the extremes of income and/or education inequality.[38] When we include repeated cross-sectional estimates (as in Table 7), we find that the average educational achievement and the Gini index of educational inequality (in level and squared level) are statistically significant in five of eight cases, but now the non-linear relationship is of the inverted-U-shaped type (which is in line with human capital investment theory). One potential reason for this instability is that omitted variables might contribute to a reversion in the trend in income inequality. In Figure 4 we have graphed the coefficients of yearly dummies obtained in

---

[37] However, this result is not robust. When we introduce a proxy for technological progress (the capital/output ratio) in the regressions (see Table A2 in Appendix 2), the inequality in educational attainment retains its sign and significance even with the hyperbolic functional form. Notice that the number of observations is reduced under this specification because we do not have information about national capital stocks for 18 of the countries.

[38] Regional dummies (used in the estimates of random effects reported in Table A2) indicate that the greatest inequality was registered in Latin America and sub-Saharan Africa, where inequality indexes were 6 percentage points higher than they were in the OECD countries (which represent the reference case; see the fifth column in Table A2). Conversely, the distribution was more egalitarian in the currently (or previously) centrally planned economies, where the Gini index was 12 percentage points lower than it was in the OECD, and in South Asia, North Africa and the Middle East.

the regressions reported in the fifth columns of Tables 5 and A2. These coefficients (normalized by the coefficient of the initial year) measure a shift in the intercepts of the regressions, thus capturing part of the variance that is left unexplained by the estimated model and that is year specific. For the first half of the sample (until 1975), we witness a growing pressure for the compression of income distribution (on the order of 1 point in the Gini index every five years), whereas this effect disappears during the 1980s. In the 1990s the phenomenon works in the opposite direction, favouring widening income disparity.

Since by definition yearly/regional dummies capture unexplained components, we do not have reliable explanations for these effects that do not refer to per capita income or educational achievement. Nevertheless, we have experimented with two additional variables that may capture some of the differences among countries or years. The first one is the physical capital/output ratio. On theoretical grounds, if physical and human capital are substitutes in the aggregate production function, an increase in the former raises the productivity of the latter. Therefore, *cæteris paribus*, we will obtain higher returns to education whenever physical capital accumulation becomes more intensive. Thus, we can expect greater income inequality whenever and wherever there is intensive investment in physical capital.[39] This variable is introduced in Table 7 and also in Table A3 in Appendix 2 (which reproduces information in Table 5, though the number of observations is reduced because of missing information). This variable is not very significant in the fixed effect estimates, but has a positive and significant sign in the repeated cross-sectional estimates (up to 1985). Other things being constant, countries characterized by higher accumulation in physical capital also exhibit higher income inequality: passing from an average k/y ratio of 2 in South Asia to 3 in the OECD countries raises the Gini index of income inequality by 2 (up to 5) points. However it is insignificant in more recent years.

The second variable we take into account is the amount of public resources invested in education. If the technology for human capital formation includes invested resources, we can expect increased human capital per unit of time spent in school whenever education expenditure is raised. The resources invested in education should include both public and private expenditure for the management of educational institutions. In the absence of reliable information about private expenditure, we can use the ratio of government educational expenditure to gross domestic product. An undesirable feature of introducing new controls is the increase in the number of missing

[39] This claim is objectionable when we think of information and telecommunications technologies, for which the capital/output ratio is actually lower than it is for manufacturing, notwithstanding the fact that the earnings differentials are higher.

observations. In the first column of Table A4 in Appendix 2, we have reproduced the fifth column of Table 5 to facilitate comparison. Using the same specification, we restrict the number of cases to applicable observations for the capital/output ratio (second column), and then we introduce the capital/output ratio (third column). We observe that an increase in capital accumulation raises income inequality (though with an elasticity which is quite low); all the other variables preserve their signs and significance. We now proceed to consider the ratio of (current+capital) government expenditure on education to gross domestic product (variable $edgvsh$).[40] The fourth column reduces the sample to country/year observations corresponding to non-missing values for the $edgvsh$ variable, whereas the fifth column introduces the $edgvsh$ variable; the k/y variable is dropped in the sixth column, which makes full use of the available sample. Even in this case, we observe that countries characterized by higher public expenditure on education exhibit higher income inequality. It is obvious that countries with higher educational achievements spend more on education. However, given the fact that we are controlling for average educational achievement (variable h_c) and the distribution of educational achievement (variable $Gini_{educ}$), the additional effect could be taken as evidence that the 'quality' of human capital incorporated in the same number of years of

[40] This variable is taken from UNESCO (1998). It is missing for 1960 and 1965, and there is a sample mean of 4.25% (standard deviation: 1.86).

schooling is higher, thus generating more dispersion in earnings. In this specification, however, the capital/output ratio loses significance.[41]

[41] A third aspect that we would have liked to consider is the possibility of different returns for different educational levels, which is invoked by Gottschalk and Smeeding (1997) as one potential explanation for rising earnings inequality in the US. We know from the literature (Psacharopoulos, 1994) that returns to education differ from country to country and tend to decline with a rising level of development. But we do not have time-series proxies for this differential effect, and we are forced to leave this effect out.

Table 5 – Estimates of income inequality – 1960-1995 –
94 countries – fixed effects
(t-statistics in parentheses)

# countries:	94	94	94	94	94	94
# obs :	454	454	454	454	454	454
Depvar:	gini	gini	gini	gini	gini	gini
intcpt	46.953	47.401	49.283	59.164	57.491	48.163
	(29.91)	*(31.75)*	*(15.76)*	*(12.17)*	*(11.67)*	*(15.26)*
gdp	-0.000	-0.001			-0.000	-0.001
	(-0.77)	*(-2.39)*			*(-1.86)*	*(-2.64)*
gdp^2	-0.000					
	(-0.16)					
1/gdp		-423.050				
		(-0.23)				
ginied			-0.182	-0.310	-0.279	-0.069
			(-1.45)	*(-2.31)*	*(-2.08)*	*(-0.53)*
$ginied^2$			0.002	0.002	0.002	0.000
			(1.48)	*(1.95)*	*(2.03)*	*(0.32)*
h_c				-1.470	-1.134	
				(-2.64)	*(-1.94)*	
$1/h_c$						2.364
						(2.84)
Years	Yes	Yes	Yes	Yes	Yes	Yes
R^2(within)	0.066	0.066	0.056	0.075	0.084	0.095

Table 6 – Estimates of income inequality by world regions – 1960-1995 – fixed effects

(t-statistics in parentheses)

Model :	OECD	NthAfric	SaharAfr	SthAsia	EastAsia	LtAmerica	PlnEcon	World
# countr:	21	8	21	5	10	22	7	94
# obs :	133	25	57	31	65	117	26	454
Depvar:	gini	gini	gini	gini	gini	gini	gini	gini
intcpt	65.654	-60.263	6.306	59.959	45.058	39.637	79.638	57.491
	(8.44)	(-0.50)	(0.16)	(1.05)	(3.59)	(2.59)	(1.83)	(11.67)
gdp	0.000	-0.002	0.003	-0.003	0.000	0.000	0.003	0.000
	(0.39)	(-0.61)	(0.70)	(-0.24)	(-0.47)	(-0.26)	(1.75)	(-1.86)
ginied	-0.941	1.638	1.392	-0.716	0.269	0.810	-1.370	-0.279
	(-3.22)	(0.69)	(1.46)	(-0.67)	(0.83)	(1.41)	(-1.16)	(-2.08)
ginied²	0.011	-0.005	-0.010	0.006	-0.004	-0.008	0.010	0.002
	(2.16)	(-0.36)	(-1.28)	(0.91)	(-1.43)	(-1.25)	(0.84)	(2.03)
h_c	-1.787	4.647	-1.717	0.734	-0.894	-2.272	-4.744	-1.134
	(-1.95)	(0.58)	(-0.31)	(0.14)	(-0.77)	(-1.66)	(-1.24)	(-1.94)
Years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2(within)	0.332	0.735	0.387	0.436	0.238	0.142	0.711	0.084

Table 7 – Estimates of income inequality – yearly cross sections – robust estimates

(t-statistics in parentheses)

Year:	1960	1965	1970	1975	1980	1985	1990	1995
#countries:	40	47	60	53	55	57	65	24
Depvar:	gini	gini	gini	gini	gini	gini	gini	gini
intcpt	46.943	40.760	32.317	52.368	37.297	37.560	49.381	19.402
	(4.49)	*(3.73)*	*(2.91)*	*(6.26)*	*(3.24)*	*(3.53)*	*(2.94)*	*(0.64)*
k/y	2.294	3.595	4.943	2.358	3.926	2.507	0.795	-1.497
	(2.28)	*(2.56)*	*(4.94)*	*(4.22)*	*(3.48)*	*(2.30)*	*(0.83)*	*(-0.61)*
gdp	0.000	0.000	0.001	0.000	-0.001	-0.001	-0.001	-0.001
	(0.05)	*(0.28)*	*(1.22)*	*(-0.56)*	*(-1.98)*	*(-2.08)*	*(-1.82)*	*(-1.97)*
ginied	0.118	0.194	0.574	-0.012	0.291	0.294	0.375	1.150
	(0.56)	*(0.60)*	*(1.92)*	*(-0.05)*	*(1.08)*	*(1.15)*	*(0.94)*	*(2.34)*
ginied²	-0.002	-0.002	-0.006	-0.001	-0.004	-0.003	-0.006	-0.013
	(-0.94)	*(-0.76)*	*(-2.26)*	*(-0.50)*	*(-1.72)*	*(-1.39)*	*(-1.49)*	*(-2.44)*
h_c	-1.457	-1.980	-2.967	-2.493	-1.052	-0.719	-1.532	2.275
	(-0.98)	*(-1.47)*	*(-3.02)*	*(-3.09)*	*(-1.20)*	*(-0.80)*	*(-1.28)*	*(0.67)*
R^2	0.263	0.23	0.568	0.498	0.538	0.45	0.38	0.446

Summing up, we have found that per capita income and average years of education in the population negatively affect income inequality. Some additional explanatory contribution is provided by the distribution of educational attainments in the population, and this variable exhibits a non-linear relationship with income inequality. Higher investment in physical capital (as proxied by capital/output ratio) or in human capital formation (as proxied by the ratio of educational expenditure to gross output) contributes to higher income inequality. These results are robust to alternative specifications, and we therefore go back to our initial (and preferred) specification, which is provided in the fifth column of Table 5 and reproduced here for simplicity (yearly dummies not shown):

$$Gini_{\text{income}} = \underset{(11.6)}{57.49} - \underset{(1.86)}{0.004} \cdot gdp - \underset{(2.08)}{0.279} \cdot Gini_{educ} + \underset{(2.03)}{0.002} \cdot Gini_{educ}^2 - \underset{(1.94)}{1.13} \cdot \overline{HC} \qquad (11)$$

If we take into account that, on the same sample, fixed effect regression yields (again, yearly dummies are not shown here):

$$Gini_{\text{educ}} = \underset{(43.3)}{71.37} - \underset{(22.84)}{6.77} \cdot \overline{HC} \qquad (12)$$

and we replace equation (12) into equation (11), we get

$$Gini_{\text{income}} = 37.72 - 0.004 \cdot gdp - 1.18 \cdot \overline{HC} + 0.091 \cdot \overline{HC^2} \qquad (13)$$

Equation (13) tells us that, *for a given level of per capita income*, income inequality has a U-shaped relationship with the average years of education in the population, with a turning point around 6.48 years. For all countries below this threshold, the two variables are negatively correlated, while the two become positively correlated above this threshold. Using the regional averages reported in Table 3, we can say that additional education promotes inequality in the OECD countries (and very recently also in the formerly planned economies), whereas it is beneficial with respect to inequality in the other regions of the world.

A final perspective on the relevance of educational achievement in predicting income inequality can be obtained by manipulating equation (10), which can be rearranged as:

$$\frac{\alpha}{\alpha+\beta\overline{\overline{HC}}}=1-\frac{G_{\text{log-income}}}{G_{ed}} \tag{14}$$

Equation (14) tells us that 1 minus the ratio between the inequality in (log)incomes and the inequality in education can provide a rough estimate of the ratio between the income of an uneducated person and the income of a person with average education. The problem is that we do not have information on either individual earnings (or incomes), and we therefore cannot compute the Gini index of logarithms of these variables, as required in equation (14). However, using simulations based on the observed distribution of educational achievement in the sample, we have computed the Gini index on both incomes and log-incomes. The two measures are proportionally related, with the goodness of the fit declining with the rate of return, β, assumed in the simulation.[42] Using this result, we have computed an (estimated) Gini index of log-income that allows us to obtain the measure proposed in equation (14). This is depicted in Figure 5. From the dynamics of this indicator at regional level, we notice that the educational premium is higher in the OECD countries (mainly because they have a higher average educational achievement), followed by Asia and Latin America. In all cases but one, this premium has been declining in recent years. In contrast, the return to education seems to be rising in the formerly planned economies.

5. Conclusions

Our plan in this paper has been to measure the inequality in educational achievement by constructing a Gini index of educational attainment. We have then used the proposed measure to analyse the relationship between inequality in incomes and inequality in educational achievement (in terms of both the average attainments and the concentration of educational achievement). Though theoretical considerations based on the theory of human capital investment suggest that

---

[42] For example, the estimated equation, assuming $\alpha = 100$ and $\beta = 0.1$, is

$$Gini_{\text{log-income}} = -\underset{(2.96)}{0.005} + \underset{(123.8)}{0.009} \cdot Gini_{\text{income}}, \quad R^2(within) = 0.95, \quad obs = 848$$

Based on an average among several simulations obtained by varying α, or β, or both, we have computed a measure of the Gini index of log(income). However, since the right-hand variable includes total incomes (and not merely earnings, as the pure theory of human capital would require), the estimated measure of log-incomes is only an approximation of what we would have liked to measure to evaluate the ratio uneducated / educated.

we should expect a non-linear relationship between these two measures of inequality, we have seen that the actual data indicate that average years of education have a stronger negative impact on measured income inequality. Multivariate regressions also demonstrate that, if we take into account the negative correlation between average educational achievement and the dispersion of educational achievement, the relationship between income inequality and average years of schooling is U-shaped, with a lower turning point at 6.5 years. Obviously, income inequality is also negatively related to per capita income; other things being constant, countries characterized by higher accumulation or greater government expenditure on education experience higher income inequality.

But most of these results are region specific. Even if the world has experienced what can called an ‘educational cycle’ during the post-war period (an increase in the average schooling by 2.2 years and a reduction in the Gini index of educational inequality by about 9), the consequences on income inequality were not univocal. Looking at the relationship between education and income inequality at regional level (as in Figure 6, that depicts the weighted mean values for each time-unit of observation), we notice that at least three separate patterns can be identified in the ‘educational cycle’ at world level. North Africa and South Asia exhibit the first pattern. Most of the countries in these regions started from a quite low initial base of educational attainment (around one year of average schooling in North Africa and South Asia in the 1960s), but were quite effective in improving the situation, more than quadrupling this average. These are not the only regions in which we find that education has the effect of reducing inequality.

A second pattern is represented by East Asia and sub-Saharan Africa, which initially followed the first pattern, though at a slower speed (the average years of schooling increased from 3.7 to 6.4 and from 1.0 to 2.7, respectively, during 1960-95). The 'leap forward' in educational attainment in these countries seems to have been insufficient to modify basic social structures (in contrast to the successful countries in the first group). Inequality in education initially declined, but after the 1970s there was a trend reversal, and this was accompanied by an increase in income inequality.

Finally, the third group is formed by the Latin American countries and the (formerly or currently) centrally planned economies. Both sets of countries were characterized by high initial levels of education (3.1 and 3.9 years on average, respectively, in 1960); nonetheless, they were able to raise the average significantly (to 6.2 and 8.2 years, respectively, by 1995). Educational inequality declined, but income inequality rose substantially, as indicated by the Gini index: 6 additional points in Latin America and more than 10 points in the planned economies.

The OECD countries represent a story on their own. The only group with average educational attainments above the threshold of 6.5 years, these countries experienced a widening in educational differentials during the entire sample period that was accompanied, after 1975, by rising income inequality.

A general lesson emerges from this evidence: increased access to education reduces income inequality only if two conditions are met. First, the initial level of educational attainment must be sufficiently low; second, the average educational attainment must be raised sufficiently rapidly. A potential explanation of these results is offered by the interaction between the supply and the demand of human capital, that is, the educational choices of the population and job creation by firms.[43] When the average educational level in the population is low, there are very few highly educated people who are likely to obtain high salaries. At the same time, there are no incentives for the creation of new jobs for skilled workers since firms are constrained by factor supply. However, when more and more educated people begin entering the labour market, the speed of technological innovation increases, followed by the creation of more skilled jobs. More people earn higher wages, and as a consequence income inequality starts declining. When the bulk of the labour force has at least a primary level of education,

[43] On the relationship between the availability of skills and job creation, see Agemoglu (1995, 1996).

leaps in technology (such as in information technology and telecommunications) are possible because skilled workers can now accomplish the more sophisticated tasks. The rise in the productivity of these workers is reflected in their remuneration, thus inducing a trend reversal in income inequality. In this way, we replicate the non-linear relationship between average educational attainment and income inequality, which is also conditioned by the level of technical development.

One may object that causality can work in the opposite direction: lower income inequality facilitates access to education and therefore contributes to a reduction in the inequality in education. However, this may be true only in the steady state. Thus, Checchi (1999) has shown that income inequality reduces enrolment rates, mainly at the secondary level. But enrolment rates reflect the rate of change of the existing human capital stock and therefore affect the *rate of change* in educational inequality. Yet, enrolment rates cannot affect the rate of change and the level of the same variable at the same time. In our framework, current income inequality affects *future* educational inequality, which, according to human capital theory, will shape *future* income inequality. Therefore, reverse causation may apply only along the intertemporal dimension.[44]

[44] In order to control for potential endogeneity, we have also considered the possibility of lagging the educational variables (see table A5 in Appendix 2), but despite of the reduction in the sample these variables retain their significance.

References

Agemoglu, D. 1995. Reward structures and the allocation of talent, *European Economic Review* 1: 17-34

Agemoglu, D. 1996. Changes in unemployment and wage inequality: an alternative theory and some evidence, CEPR Wp.1459

Aghion, P., E.Caroli and C.Gracia-Peñalosa. 1999. Inequality and economic growth: The perspective of new growth theories. *Journal of Economic Literature* 37 / 4: 1615-60.

Anand, S., and S.Kanbur. 1993. The Kuznets process and the inequality-development relationship. *Journal of Economic Development* 40: 25-52.

Atkinson, A.B. 1997. Measurement of trends in poverty and in the income distribution. The Microsimulation Unit, DAE wp #9712.

Atkinson, A.B. 1999. Is rising inequality inevitable ? A critique of the transatlantic consensus. Wider Annual lectures 3, November.

Atkinson, A.B., and A.Brandolini. 1999. Promise and pitfalls in the use of secondary data-set: a case study of OECD income inequality. mimeo.

Banerjee, A., and A.Newman. 1993. Occupational choice and the process of development. *Journal of Political Economy* 101(2): 274-298.

Bardone, L., M.Gittleman and M.Keese. 1998. Causes and consequences of earnings inequality in OECD countries. *Lavoro e Relazioni Industriali* 2: 13-60.

Barro, R. 1999. Inequality, growth and investment. NBER Working Paper n.7038.

Barro, R., and J.W.Lee. 1993. International comparisons of educational attainment. *Journal of Monetary Economics* 32 (3): 363-94.

Barro, R., and J.W.Lee. 1994. Data set for a panel of 138 countries. Washington: World Bank.

Barro, R., and J.W.Lee. 1996. International measures of schooling years and schooling quality. *American Economic Review*, Papers and Proceedings 86 (2): 218-23.

Barro, R., and J.W.Lee. 1997. Schooling quality in a cross section of countries. NBER Working Paper 6198, September.

Becker, G. 1993. *Human capital: a theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press (first edition 1964)

Benabou, R. 1996a. Equity and efficiency in human capital investment: the local connection. *Review of Economic Studies* 63: 237-264.

Benabou, R. 1996b. Inequality and growth. In *NBER Macroeconomic Annual 1996*. B.Bernanke and J.Rotemberg (eds). Cambridge Mass: The MIT Press.

Borjas, G. and V.Ramey 1995. Foreign competition, market power and wage inequality. *Quarterly Journal of Economics*: 1075-1110.

Bourguignon, F. 1994. Growth, distribution and human resources. In *En route to modern growth*. G.Ranis (ed). Baltimore: John Hopkins University Press.

Bourguignon, F. 1996. Equity and economic growth: permanent questions and changing answers ? Delta Document de travail n.96-15.

Brandolini, A. 1998. A bird's-eye view of long-run changes in income inequality. mimeo

Brandolini, A., and N.Rossi. 1998. Income distribution and growth in industrial countries. In *Income Distribution and High-Quality Growth*. V.Tanzi and K.Chu (eds). Cambridge Mass: The MIT Press.

Breen, R. and C.García-Peñalosa. 1999. Income inequality and macroeconomic volatility: an empirical investigation. mimeo.

Castello, A. and R.Doménech. 2001. Human capital inequality and economic growth: some new evidence. mimeo.

Checchi, D. 1999. Inequality in incomes and access to education. A cross-country analysis (1960-90). Wider Working Paper n.158.

Cornia, G.A. 1999. Liberalization, globalization and income distribution. Wider Working Paper n.157.

Deininger, K., and L.Squire. 1996. A New Data Set Measuring Income Inequality. *World Bank Economic Review* 10 (3): 565-591.

Deininger, K., and L.Squire. 1998. New ways of looking at old issues: inequality and growth. *Journal of Development Economics* 57: 259-287.

Feenstra, R. and G.Hanson. 1996. Globalization, outsourcing and wage inequality. *American Economic Review* 86/2: 240-251.

Filmer, D. and L.Pritchett. 1998. The effect of household wealth on educational attainment – Demographic and health survey evidence. World Bank Policy Research working paper n.1980.

Flug, K., A.Spilimbergo and E.Wachtenheim. 1998. Investment in education: do economic volatility and credit constraints matter ? *Journal of Development Economics* 55: 465-481.

Freeman R. 1986, Demand for education, in Ashenfelter O.-Layard R. 1986 (eds), *Handbook of labour economics*, NorthHolland, New York

Galor, O. and J.Zeira. 1993. Income distribution and macroeconomics. *Review of Economic Studies* 60: 35-52.

Galor, O. and D.Tsiddon. 1996. Income distribution and growth: the Kuznets hypothesis revisited. *Journal of*

Galor, O. and D.Tsiddon. 1997. The distribution of human capital and economic growth. *Economica* 63: S103-S117.

Glyn, A. and W.Salverda. 2000. Employment inequalities. in M.Gregory and S.Bazen (eds). *Low-wage employment in Europe.* Oxford University Press, Oxford.

Gradstein, M., and B.Milanovic. 2000. Does liberté=egalité ? A survey of the empirical evidence on the links between political democracy and income inequality, mimeo.

Gottschalk, P, and T.Smeeding. 1997. Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature* 35 (June): 633-687.

Higgins, M. and J.Williamson. 1999. Explaining inequality the world round: cohort size, Kuznets curve, and openness. mimeo, June.

Inter-American Development Bank. 1998. *Facing up to inequality in Latin America – 1998-99 Report*. Baltimore: John Hopkins University Press.

Justman, M., and M.Gradstein. 1999, The democratization of political elites and the decline in inequality in modern economic growth. in E.Brezia and P.Temin (eds). *Elites, minorities, and economic growth*. Elsevier, Amsterdam.

Kahn, L. 2000. Wage inequality, collective bargaining and relative employment 1985-94: evidence form 15 OECD countries. forthcoming in *Review of Economics and Statistics*.

Kuznets, S. 1955. Economic growth and income inequality. *American Economic Review* 45: 1-28.

Li, H., L.Squire and H.Zou. 1998, Explaining international and intertemporal variations in income inequality. *Economic Journal* 108: 26-43.

Londoño, J.L. 1996. *Inequality and poverty in Latin America during the last four decades*. Washington: World Bank Latin American and Caribbean Studies.

Lopez, R., V.Thomas and Y.Wang. 1998. Addressing the education puzzle. The distribution of education and economic reform. World Bank Policy Research Working Paper n.2031.

Milanovic, B. 1999. True world income distribution, 1988 and 1993. World Bank Policy Research Working Paper n.2244.

Mincer, J. 1974. *Schooling, Experience, and Earnings,* New York: NBER.

Mincer, J. 1996. Changes in wage inequality, 1970-1990. NBER Working Paper 5823, November.

Murphy, K., C.Riddel and P.Romer. 1998. Wages, skills and Technology in the United States and Canada. mimeo.

Nehru, V., and A.Dhareshwar. 1993. A New Database on Physical Capital Stock: Sources, Methodology and Results. *Rivista de Analisis Economico* 8 (1): 37-59.

Neal, D. and S.Rosen. 1998. Theories of the distribution of labor earnings. NBER wp. #6378.

O'Neill, D. 1995. Education and income growth: implications for cross-country inequality. *Journal of Political Economy* 103(6): 1289-1299.

Perotti, R. 1993. Political equilibrium, income distribution and growth. *Review of Economic Studies* 60: 755-776.

Perotti, R. 1996. Growth, income distribution and democracy: what the data say. *Journal of Economic Growth* 1(2): 149-187.

Piketty, T. 1997. The dynamics of the wealth distribution and the interest rate with credit rationing. *Review of Economic Studies* 64(2): 173-189.

Psacharopoulos, G. 1994. Returns to Investment in Education: A Global Update. *World Development* 22(9): 1325-1343.

Ram, R. 1990. Educational expansion and schooling inequality: international evidence and some implication. *Review of Economics and Statistics* :266-273.

Ravallion, M. and Q.Wodon. 1999. Does child labor displace schooling ? Evidence on behavioral responses to an enrolment subsidy. World Bank Policy Research wp #2116.

Skyt Nielsen, H. 1999. Child labor and school attendance: two joint decisions. mimeo.

Summers, R. and A.Heston. 1991. The Penn World table (Mark 5): an expanded set of international comparisons, 1950-1988. *Quarterly Journal of Economics* 56(2): 327-368.

UNESCO 1998. *Statistical yearbook.* New York.

Vinod, T., W.Yang and F.Xibo. 2001. Measuring education inequality – Gini coefficients of education. World Bank. Policy Research Working Paper #2525.

Weiner, M. 1991. *The child and the State in India.* New Delhi: Oxford University Press.

World Bank 1998. *World Bank Data on CD-ROM*. Washington DC.

Appendix 1 – Data source

We have taken seriously the recommendation of Atkinson and Brandolini (1999). Data on income inequality are mainly from Deininger and Squire (1996)[45]. In addition, 12 observations (average = 35.05) referred to OECD countries are from Brandolini (1998) and 25 (average = 43.54) are from World Bank (1998). Finally 7 observations (average = 37.65) are from WIID (World Income Inequality Dataset, downloadable at http://www.wider.unu.edu/wiid/wiid.htm. On the whole we have 546 observations referred to 113 countries (with an average of 4.8 observations per country)[46].

While there are no significant differences in Gini indexes when the recipient unit is the (equivalised) household or the individual, we find an average difference of 6.47 percentage points when the same measure is based on gross incomes instead of net ones.[47] We could have introduced a dummy variable controlling for the income definition (as in Deininger and Squire 1998), but in such a case we would have dispensed with all observations for which this information was absent. For this reason, we have preferred to augment the measures based on net incomes by the average difference.[48]

---

[45] Downloaded on 22 October 1998. Among these, 349 observations are labelled 'high quality' (average = 38.79), and 153 observations are labelled 'low quality' (average = 45.87).

[46] The number of observations reduces to 471 (corresponding to 97 countries with an average of 4.9 observations per country) when we restrict to the cases with non-missing observations on educational variables.

[47] By regressing Gini index on income distribution on a dummy variable INCOME (which is equal to 1 when the recipient unit is the equivalised household, and 0 when is the individual) we get

$$Gini = \underset{(58.9)}{41.78} - \underset{(1.03)}{1.01} \cdot INCOME, \qquad R^2 = 0.00, \qquad n = 471$$

On the contrary, creating a dummy variable TYPE (equal to 1 when inequality measure is based on gross incomes, and 0 when is based on net income) we get

$$Gini = \underset{(35.9)}{35.94} + \underset{(6.46)}{6.46} \cdot TYPE, \qquad R^2 = 0.10, \qquad n = 369\,.$$

[48] Similar correction was applied to Gini measures based on rural samples (5 observations), which on average resulted higher than national coverage samples by 8.94 points.

Data on physical capital stocks are from Nehru and Dhareshwar (1993). Data on per-capita income and educational achievements are from Barro and Lee (1993, 1994, 1996, and 1997).[49] In particular, data on estimated length of schooling $n_i, i = p,s,h$, have been obtained by dividing the average years of schooling for a given level of education by the population share who completed that level of education: using Barro and Lee (1996) definitions:[50]

$$n_p = \frac{\text{pyr25}}{\text{pri25} + \text{sec25} + \text{high25}}, \quad n_s = \frac{\text{syr25}}{\text{sec25} + \text{high25}}, \quad n_h = \frac{\text{hyr25}}{\text{high25}}.$$

Where possible, the series have been updated to 1995 using World Bank (1998) and UNESCO (1998). Data on average years of schooling for 1995 have been estimated based on the corresponding enrolment rates for the previous three decades.

The list of 97 countries for which we have non-missing observations on inequality in incomes and inequality in educational achievements is as follows (the number of available observations is given in brackets):

Sub-Saharan Africa:
Botswana (3), Cameroon (1), Central African Republic (1), Gambia (1), Ghana (3), Guinea-Bissau (1), Kenya (7), Lesotho (1), Liberia (1), Malawi (4), Mauritius (3), Niger (1), Rwanda (1), Senegal (3), Sierra Leone (3), South Africa (6), Sudan (2), Tanzania (6), Uganda (3), Zambia (4), Zimbabwe (2).
North Africa and Middle East:
Algeria (2), Egypt (3), Tunisia (7), Iran (3), Israel (5), Jordan (3), North Yemen (1), Cyprus (1).
East Asia and the Pacific:
Hong Kong (7), Indonesia (7), Japan (7), Korea (7), Malaysia (7), Philippines (7), Singapore (6), Taiwan (7), Thailand (7), Fiji (3).
South Asia:
Bangladesh (7), India (7), Nepal (3), Pakistan (7), Sri Lanka (7).
Latin America and the Caribbean:
Barbados (4), Reunion (1), Costa Rica (8), Dominica (4), El Salvador (6), Guatemala (4), Honduras (4), Jamaica (7), Mexico (8), Nicaragua (1), Panama (6), Trinidad and Tobago (5), Argentina (6), Bolivia (3), Brazil (7), Chile (7), Colombia (8), Ecuador (4), Guyana (2), Paraguay (3), Peru (6), Uruguay (7), Venezuela (7).

[49] Barro and Lee (1994) is in turn based on Summers and Heston (1991).

[50] This procedure yields unreasonable values of n_p for few observations. In these cases, these values have been replaced either with the corresponding values computed on the population with more than 15 years or with the legal duration of primary education (as measured in 1965 – variable *durp* in the original Barro-Lee data set).

OECD countries:
Australia (8), Austria (4), Belgium (6), Canada (8), Denmark (6), Finland (8), France (8), (West) Germany (8), Greece (6), Ireland (5), Italy (6), Netherlands (7), New Zealand (7), Norway (7), Portugal (3), Spain (6), Sweden (7), Switzerland (2), Turkey (6), United Kingdom (8), United States (8).
Centrally planned economies:
China (4), Cuba (3), Czechoslovakia (7), Hungary (7), Yugoslavia (6), Bulgaria (7), Romania (1), (former) Soviet Union (5).

Appendix 2 – Additional tables

Table A1 – Probit estimates of data availability – 1960-1995 – 139 countries

(t-statistics in parentheses)

# obs :	992	992	992
Depvar:	income	education	inc&educ
intcpt	-7.961	-3.966	-8.302
	(-8.84)	*(-4.09)*	*(-8.88)*
year=65	-0.153	0.065	0.101
	(-0.81)	*(0.32)*	*(0.51)*
year=70	0.233	0.078	0.428
	(1.23)	*(0.38)*	*(2.15)*
year=75	-0.120	0.279	0.162
	(-0.62)	*(1.31)*	*(0.81)*
year=80	-0.174	0.269	0.016
	(-0.90)	*(1.24)*	*(0.08)*
year=85	-0.233	0.133	0.053
	(-1.23)	*(0.65)*	*(0.27)*
year=90	0.090	0.056	0.303
	(0.48)	*(0.27)*	*(1.54)*
year=95	-1.196	-0.444	-1.414
	(-4.73)	*(-1.69)*	*(-5.19)*
northafric	-0.772	-0.402	-0.961
middleast	*(-3.83)*	*(-1.69)*	*(-4.57)*
subsaharan	0.025	-0.468	-0.250
africa	*(0.10)*	*(-1.64)*	*(-0.99)*
southasia	0.160	0.742	0.283
	(0.53)	*(1.54)*	*(0.92)*
eastasia	0.420	0.061	0.521
pacific	*(1.87)*	*(0.22)*	*(2.27)*
latin	0.505	0.644	0.577
america	*(2.63)*	*(2.48)*	*(2.93)*
centr.pl.	0.726	-1.327	-0.194
economies	*(2.11)*	*(-4.26)*	*(-0.65)*
log(pop)	0.439	0.272	0.413
	(12.01)	*(7.38)*	*(11.18)*
log(gdp)	0.537	0.357	0.576
	(6.27)	*(3.80)*	*(6.46)*
growth	3.371	-4.574	3.178
populat	*(1.77)*	*(-2.49)*	*(1.47)*
R^2	0.287	0.253	0.341

Table A2– Estimates of income inequality – 1960-1995 – 94 countries – random effects

(t-statistics in parentheses)

# countries:	94	94	94	94	94	94
# obs :	454	454	454	454	454	454
Depvar:	gini	gini	gini	gini	gini	gini
intcpt	46.389	45.999	38.889	53.066	53.149	43.112
	(18.96)	*(21.75)*	*(18.11)*	*(12.16)*	*(12.21)*	*(15.83)*
gdp	-0.001	-0.001			0.000	-0.001
	(-1.84)	*(-3.65)*			*(-2.13)*	*(-3.57)*
gdp^2	0.000					
	(0.43)					
1/gdp		-425.957				
		(-0.30)				
ginied			0.102	-0.108	-0.096	0.135
			(1.07)	*(-0.99)*	*(-0.88)*	*(1.35)*
$ginied^2$			-0.001	0.000	0.001	-0.001
			(-0.56)	*(0.46)*	*(0.57)*	*(-1.39)*
h_c				-1.557	-1.136	
				(-3.71)	*(-2.46)*	
$1/h_c$						2.464
						(3.19)
northafric	-2.161	-1.850	0.160	-1.118	-3.162	-2.840
middleast	*(-0.73)*	*(-0.64)*	*(0.05)*	*(-0.38)*	*(-1.03)*	*(-0.92)*
subsaharan	7.381	8.222	11.021	7.864	5.541	6.154
africa	*(2.69)*	*(2.94)*	*(4.53)*	*(3.08)*	*(1.99)*	*(2.23)*
southasia	-5.373	-4.606	-1.991	-4.236	-6.749	-7.090
	(-1.52)	*(-1.32)*	*(-0.60)*	*(-1.28)*	*(-1.91)*	*(-2.01)*
eastasia	-0.588	-0.287	1.953	1.130	-0.580	-1.298
pacific	*(-0.23)*	*(-0.11)*	*(0.83)*	*(0.48)*	*(-0.23)*	*(-0.52)*
latin	7.135	7.411	10.291	8.069	6.338	6.340
america	*(3.19)*	*(3.41)*	*(5.26)*	*(3.98)*	*(2.89)*	*(2.91)*
centr.pl.	-13.562	-13.367	-9.255	-9.537	-12.206	-13.339
economies	*(-4.52)*	*(-4.50)*	*(-3.46)*	*(-3.61)*	*(-4.16)*	*(-4.62)*
Years	Yes	Yes	Yes	Yes	Yes	Yes
R^2(overall)	0.52	0.52	0.50	0.52	0.53	0.54

Excluded case: OECD / 1960

Table A3 – Estimates of income inequality – 1960-1995 – fixed effects - using capital / output ratio

(t-statistics in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
# countries:	76	76	76	76	76	76
# obs :	401	401	401	401	401	401
Depvar:	gini	gini	gini	gini	gini	gini
intcpt	47.109	48.409	51.909	67.357	65.054	54.084
	(22.21)	*(23.21)*	*(13.75)*	*(12.76)*	*(12.20)*	*(13.25)*
k/y	0.860	0.931	0.317	0.714	0.791	0.546
	(1.18)	*(1.30)*	*(0.44)*	*(1.00)*	*(1.12)*	*(0.76)*
gdp	-0.001	-0.001			-0.001	-0.001
	(-1.56)	*(-2.93)*			*(-2.36)*	*(-3.11)*
gdp^2	0.000					
	(0.46)					
1/gdp		-3515.234				
		(-1.82)				
ginied			-0.332	-0.556	-0.515	-0.404
			(-2.39)	*(-3.81)*	*(-3.53)*	*(-2.62)*
$ginied^2$			0.003	0.004	0.005	0.005
			(2.37)	*(3.29)*	*(3.44)*	*(2.84)*
h_c				-2.470	-2.059	
				(-4.09)	*(-3.29)*	
$1/h_c$						-3.641
						(-1.07)
Years	Yes	Yes	Yes	Yes	Yes	Yes
R^2(within)	0.089	0.098	0.079	0.125	0.141	0.114

Table A4 – Estimates of income inequality – 1960-1995 –
fixed effects – using educational expenditure

(t-statistics in parentheses)

# countries:	94	76	76	69	69	75
# obs :	454	401	401	241	241	256
Depvar:	gini	gini	gini	gini	gini	gini
intcpt	57.491	66.599	65.054	65.538	66.066	68.070
	(11.67)	*(12.93)*	*(12.20)*	*(6.69)*	*(6.82)*	*(7.20)*
gdp	0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(-1.86)	*(-2.31)*	*(-2.36)*	*(-3.25)*	*(-3.47)*	*(-2.95)*
ginied	-0.279	-0.529	-0.515	-0.614	-0.718	-0.648
	(-2.08)	*(-3.64)*	*(-3.53)*	*(-3.07)*	*(-3.53)*	*(-3.10)*
ginied^2	0.002	0.005	0.005	0.008	0.009	0.007
	(2.03)	*(3.59)*	*(3.44)*	*(4.21)*	*(4.67)*	*(3.95)*
h_c	-1.134	-1.977	-2.059	-1.419	-1.523	-1.921
	(-1.94)	*(-3.18)*	*(-3.29)*	*(-1.70)*	*(-1.85)*	*(-2.43)*
k/y			0.791	0.570	0.030	
			(1.12)	*(0.60)*	*(0.03)*	
edgvsh					0.979	0.902
					(2.22)	*(2.04)*
Years	Yes	Yes	Yes	Yes	Yes	Yes
R^2(within)	0.084	0.137	0.141	0.194	0.218	0.169

Table A5 – Estimates of income inequality – 1960-1995
– fixed effects – lagged educational variables
(t-statistics in parentheses)

#countries:	94	94	94	94	94
# obs :	432	432	432	432	432
Depvar:	gini	gini	gini	gini	gini
intcpt	48.360	49.631	50.589	58.623	57.049
	(40.23)	*(15.45)*	*(16.16)*	*(11.07)*	*(10.15)*
gdp	0.000	-0.001	-0.001	0.000	0.000
	(-2.04)	*(-2.15)*	*(-2.28)*	*(-1.43)*	*(-1.75)*
ginied		-0.098		-0.224	
		(-0.76)		(-1.58)	
$ginied^2$		0.001		0.002	
		(1.03)		(1.43)	
$ginied_{-1}$			-0.137		-0.220
			(-1.17)		(-1.67)
$ginied^2_{-1}$			0.002		0.002
			(1.48)		(1.78)
h_c				-1.328	
				(-2.13)	
$h_{c,-1}$					-0.654
					(-1.38)
Years	Yes	Yes	Yes	Yes	Yes
	---	---	---	---	---
R^2(within)	0.077	0.081	0.084	0.093	0.089

Appendix 3 – Alternative education inequality measures

Let us consider two groups in the population: one group has 0 years of education and has dimension $(1-HC)$; the second group has n years of education and has dimension equal to HC. Given the fact that the average educational attainment is $n \cdot HC$, when we can compute the standard deviation of educational attainment obtaining

$$St.Dev. = \sqrt{(1-HC)(0-n\cdot HC)^2 + HC(n-n\cdot HC)^2} = \sqrt{n\cdot\left(HC-HC^2\right)} \quad \text{(A.1)}$$

which by construction exhibits an inverted U-shaped relationship between average educational achievement and its standard deviation.[51] Educational inequality (as measured by standard deviation) will rise until the population with n years of education will have reached 0.5, and then will start declining. In fact

$$\frac{dSt.Dev.}{dHC} = \frac{\sqrt{n}}{2\sqrt{HC-HC^2}}\cdot(1-2\cdot HC) > 0 \quad iff \quad HC < \frac{1}{2} \quad \text{(A.2)}$$

If we adopt a different measure for educational inequality (like the Gini concentration index on educational attainment) for an analogous case, we get

$$Gini_{ed} = \frac{2\cdot\left[|n-0|\cdot HC\cdot(1-HC)\right]}{2\cdot n\cdot HC} = (1-HC) \quad \text{(A.3)}$$

which is obviously linear in educational attainment. When no one has education, $Gini_{ed}$ is obviously zero. When the first person goes to school, $Gini_{ed}$ jumps to 1 and then declines steadily as more and more people go to school.

Moving to income distribution, let us assume that income is solely determined by educational achievement (i.e. $\lg(y_i) = \alpha + \beta\cdot n_i$). Therefore a group $(1-HC)$ of the population will earn $[\exp(\alpha)]$, and the other complementary group will earn $[\exp(\alpha+\beta\cdot n)]$. The Gini indexes on incomes (or log-incomes) distribution are rather similar

[51] Therefore, the results obtained by Ram (1990) and Londoño (1996), where they show an inverted U-shaped relationship between average educational attainment and education inequality (as measured by the standard deviation) are almost implied by the inequality measure that has been chosen.

$$Gini_{\lg(y)} = \frac{2\cdot[|\alpha+\beta n-\alpha|\cdot HC\cdot(1-HC)]}{2\cdot[\alpha\cdot(1-HC)+\beta\cdot n\cdot HC]} = \frac{\beta\cdot n\cdot HC\cdot(1-HC)}{\alpha+\beta\cdot n\cdot HC} = \frac{HC\cdot(1-HC)}{\frac{\alpha}{\beta n}+HC} \quad \text{(A.4)}$$

$$Gini_{y} = \frac{2\cdot[|\exp(\alpha+\beta n)-\exp(\alpha)|\cdot HC\cdot(1-HC)]}{2\cdot[\exp(\alpha)\cdot(1-HC)+\exp(\alpha+\beta n)\cdot HC]} = \frac{(\exp(\beta n)-1)\cdot HC\cdot(1-HC)}{1+(\exp(\beta n)-1)\cdot HC} = \frac{HC\cdot(1-HC)}{\frac{1}{(\exp(\beta n)-1)}+HC} \quad \text{(A.5)}$$

It can be proved that there is a non-linear relationship between HC and either $Gini_{\lg(y)}$ or $Gini_y$. Computing the first derivative of $Gini_y$ with respect to HC yields

$$\frac{dGini_y}{dHC} = \frac{(\exp(\beta n)-1)^2\cdot\left[-HC^2+\frac{2}{(\exp(\beta n)-1)}\cdot HC-\frac{1}{(\exp(\beta n)-1)}\right]}{[1+(\exp(\beta n)-1)\cdot HC]^2} > 0$$

$$iff \qquad HC \in \left[0, -\frac{1}{(\exp(\beta n)-1)}+\sqrt{\frac{1}{(\exp(\beta n)-1)}\cdot\left(\frac{1}{(\exp(\beta n)-1)}+1\right)}\right] \quad \text{(A.6)}$$

Therefore inequality in incomes, as measured by Gini index, will initially increase and afterwards decreases when a growing share of the population gains access to school.

One may object that this could depend on the assumption of just two groups in the population. It can be proved that this is not the case, since the same qualitative results obtain at least with three groups of population. Let us consider the case where one group has 0 years of education and has dimension $HC_n = (1-HC_p-HC_s)$; a second group of size HC_p has n_p years of schooling; and a third group has n_s years of schooling and size equal to HC_s. The mean educational attainment is $\overline{HC} = n_p HC_p + n_s HC_s$. The standard deviation has to be redefined as

$$\begin{aligned} St.Dev. &= \sqrt{(1-HC_p-HC_s)(0-\overline{HC})^2+HC_p(n_p-\overline{HC})^2+HC_s(n_s-\overline{HC})^2} = \\ &= \sqrt{n_p^2\cdot HC_p(1-HC_p)+n_s^2\cdot HC_s(1-HC_s)-2n_pn_sHC_pHC_s} \end{aligned} \quad \text{(A.7)}$$

In order to study the evolution of education inequality measures, we assume that HC_p and HC_s vary simultaneously at the same rate, i.e. the latter is proportional to the former.[52] This is equivalent to assume that $HC_s = \lambda HC_p, \lambda \le 1$ and to take the first derivative with respect to HC_p. With standard deviation it corresponds to the case of

$$\frac{dSt.Dev.}{dHC_p} = \frac{1}{2} \cdot \frac{n_p^2 + n_s^2\lambda - 2(n_p + n_s\lambda)^2 HC_p}{\sqrt{n_p^2 \cdot HC_p(1-HC_p) + n_s^2 \cdot HC_s(1-HC_s) - 2n_p n_s HC_p HC_s}} \quad (A.8)$$

The first order derivative can be signed as follows

$$\frac{dSt.Dev.}{dHC_p} > 0 \quad iff \qquad HC_p < \frac{n_p^2 + n_s^2\lambda}{2(n_p + n_s\lambda)^2} < \frac{1}{2} \quad (A.9)$$

which is similar to the result obtained in equation (A.2). Thus educational inequality (as measured by standard deviation) increases under massive schooling up to the point when HC_p passes the threshold of ½, and the declines. Therefore, even with three groups of population we replicate the inverted U-shaped relationship between educational inequality and educational achievement.
Analogously the Gini index on educational attainment can be obtained as

$$Gini_{ed} = \frac{n_s HC_s(HC_p + HC_n) + n_p HC_p(-HC_s + HC_n)}{\overline{HC}} = \frac{n_s HC_s(1-HC_s) + n_p HC_p(1-HC_p - 2HC_s)}{\overline{HC}} \quad (A.10)$$

Using the same device previously introduced (namely that $HC_s = \lambda HC_p, \lambda \le 1$), we can proceed to study the dynamical evolution of educational inequality

$$\frac{dGini_{ed}}{dHC_p} = \frac{-n_s\lambda^2 - n_p(1-2\lambda)}{n_p + n_s\lambda} < 0 \quad (A.11)$$

Passing now to actual income

---

[52] This is obviously an approximation, since education is a sequential process, and we cannot have a simultaneous expansion of enrolment in secondary school unless we have previously observed a similar increase in primary school enrolment. However, an expansion of primary education enrolment is very likely to induce an expansion in secondary education enrolment after some time. Similar results can be obtained when assuming that both HC_p and HC_s rise at the same growth rate λ: in such a case the first derivative with respect to λ has to be taken into account.

$$Gini_y = \frac{\exp(\alpha+\beta_s n_s)HC_s(HC_p+HC_n)+\exp(\alpha+\beta_p n_p)HC_p(-HC_s+HC_n)+\exp(\alpha)HC_n(-HC_s-HC_p)}{\exp(\alpha)HC_n+\exp(\alpha+\beta_p n_p)HC_p+\exp(\alpha+\beta_s n_s)HC_s} =$$

$$= \frac{\exp(\beta_s n_s)HC_s(1-HC_s)+\exp(\beta_p n_p)HC_p(1-2HC_s-HC_p)+(1-HC_s-HC_p)(-HC_s-HC_p)}{(1-HC_p-HC_s)+\exp(\beta_p n_p)HC_p+\exp(\beta_s n_s)HC_s} =$$

$$= \frac{[\exp(\beta_s n_s)-1]HC_s(1-HC_s)+[\exp(\beta_p n_p)-1]HC_p(1-2HC_s-HC_p)}{1+[\exp(\beta_p n_p)-1]HC_p+[\exp(\beta_s n_s)-1]HC_s} =$$

$$= \frac{\hat{\beta}_s HC_s(1-HC_s)+\hat{\beta}_p HC_p(1-2HC_s-HC_p)}{1+\hat{\beta}_p HC_p+\hat{\beta}_s HC_s} = \frac{-\hat{\beta}_s HC_s^2-2\hat{\beta}_p HC_s HC_p-\hat{\beta}_p HC_p^2+\hat{\beta}_p HC_p+\hat{\beta}_s HC_s}{1+\hat{\beta}_p HC_p+\hat{\beta}_s HC_s} \tag{A.12}$$

where the $\hat{\beta}_i$ have been redefined accordingly. Assuming again that $HC_s = \lambda HC_p, \lambda \le 1$, further define

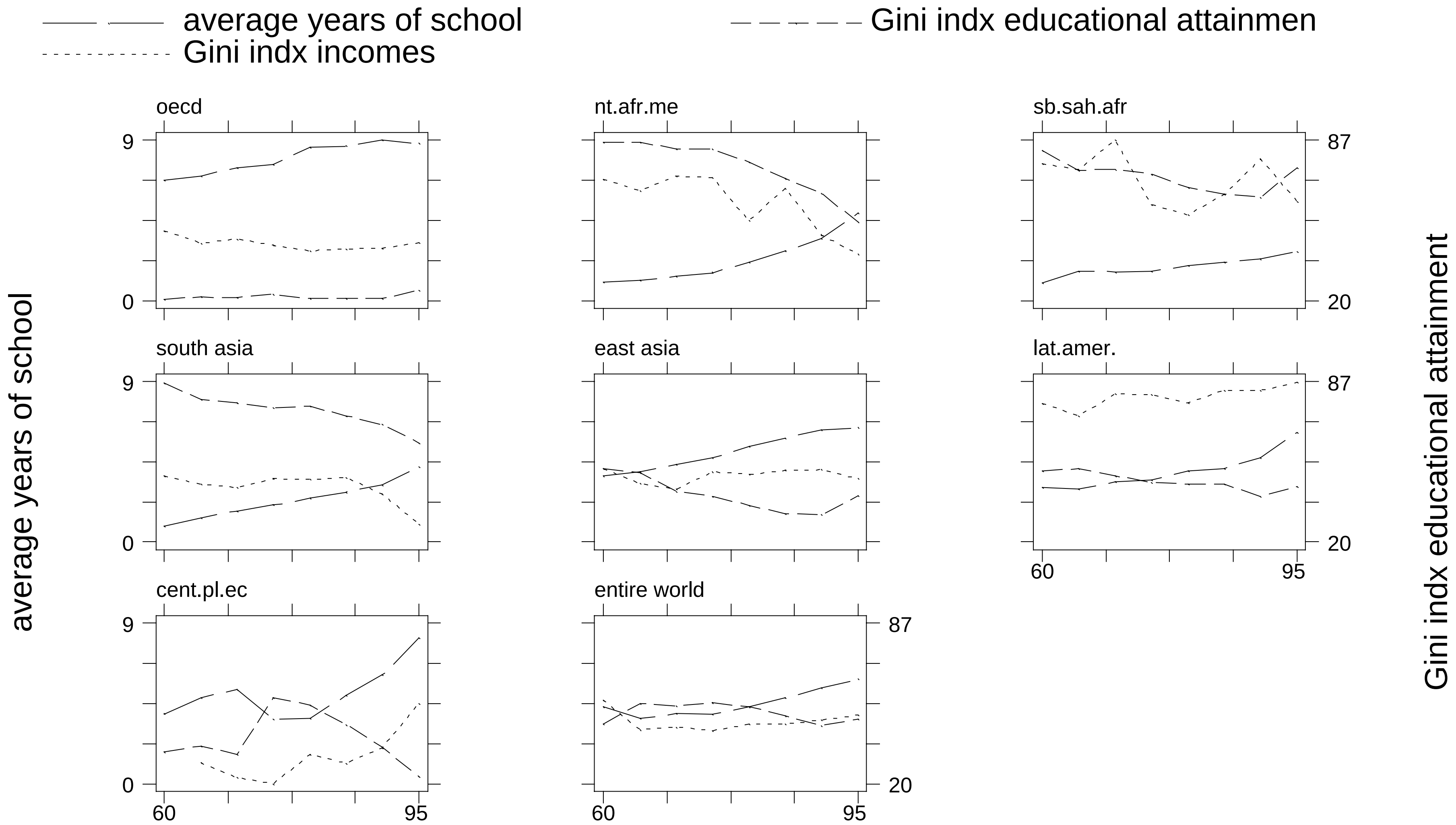
$$A = \hat{\beta}_p + \hat{\beta}_s\lambda, \;\; B = \hat{\beta}_p(1-2\lambda)+\hat{\beta}_s\lambda^2$$

Now the Gini index on income inequality appears as $Gini_y = \frac{A\cdot HC_p - B\cdot HC_p^2}{1+A\cdot HC_p}$. Thus the first order derivative with respect to HC_p is

$$\frac{dGini_y}{dHC_p} = \frac{A(1-B\cdot HC_p^2)}{(1+A\cdot HC_p)^2} \tag{A.13}$$

which is positive whenever $HC_p < \left(\hat{\beta}_p(1-2\lambda)+\hat{\beta}_s\lambda^2\right)^{-1/2}$.
As long as $\hat{\beta}_p \cong \hat{\beta}_s \Leftrightarrow \left(\hat{\beta}_p(1-2\lambda)+\hat{\beta}_s\lambda^2\right) \cong \hat{\beta}_p(1-\lambda)^2 \ge 0$. As a consequence, for low values of HC_p an increase in schooling rises income inequality; subsequently the relationship reverts to negative. This is analogous to previous result with just two subgroups of population.

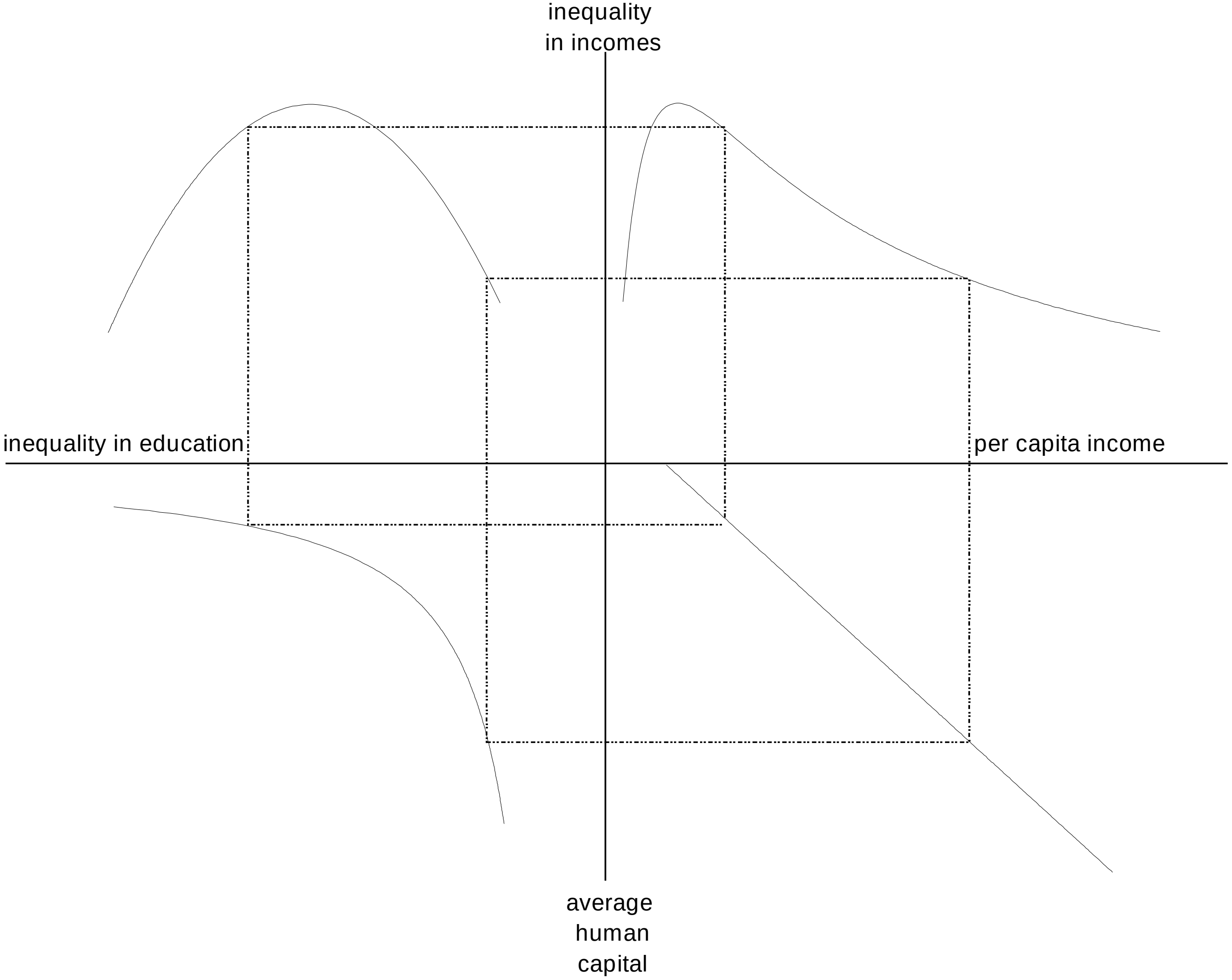
Figure 1



means - weights=population

Educational achievements and income inequality

Figure 2

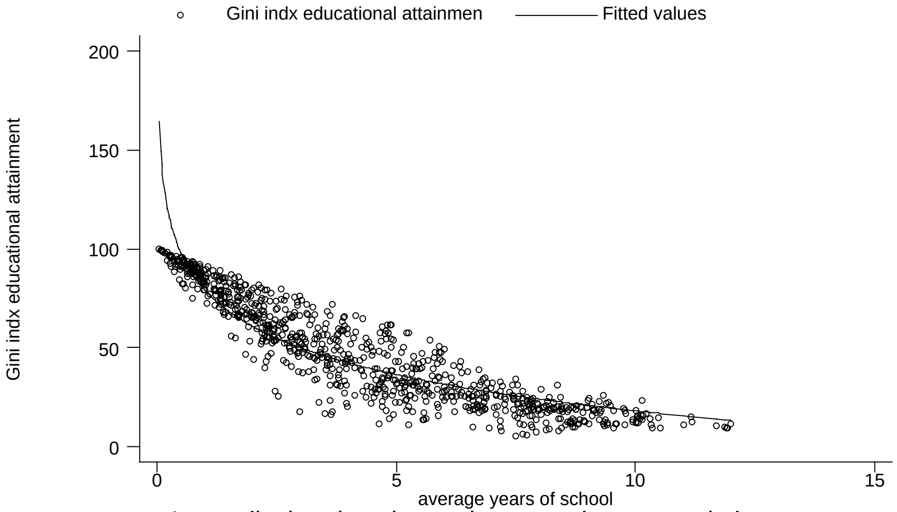


inequality
in incomes
inequality in education
per capita income
average
human
capital

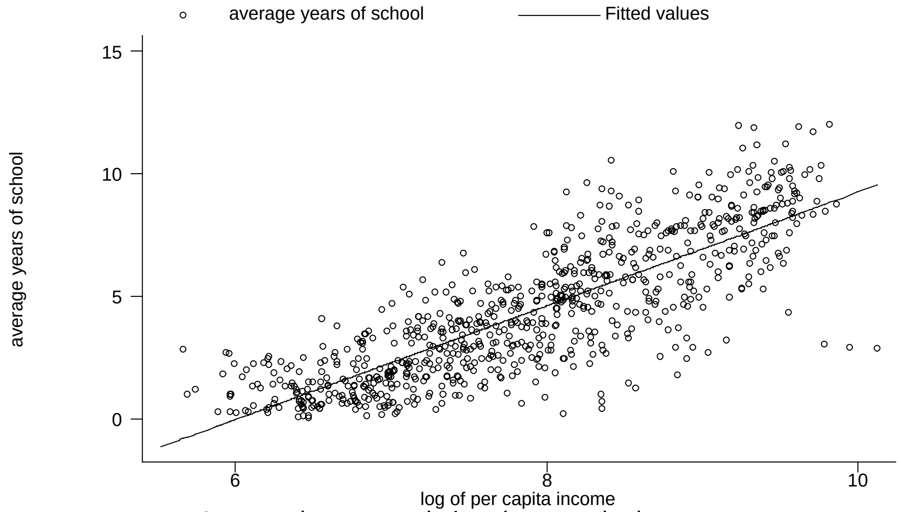
Figure 3

Inequality in incomes and inequality in education

Inequality in incomes and per-capita income



Inequality in education and average human capital



Average human capital and per-capita income

Basic relationships

Figure 4

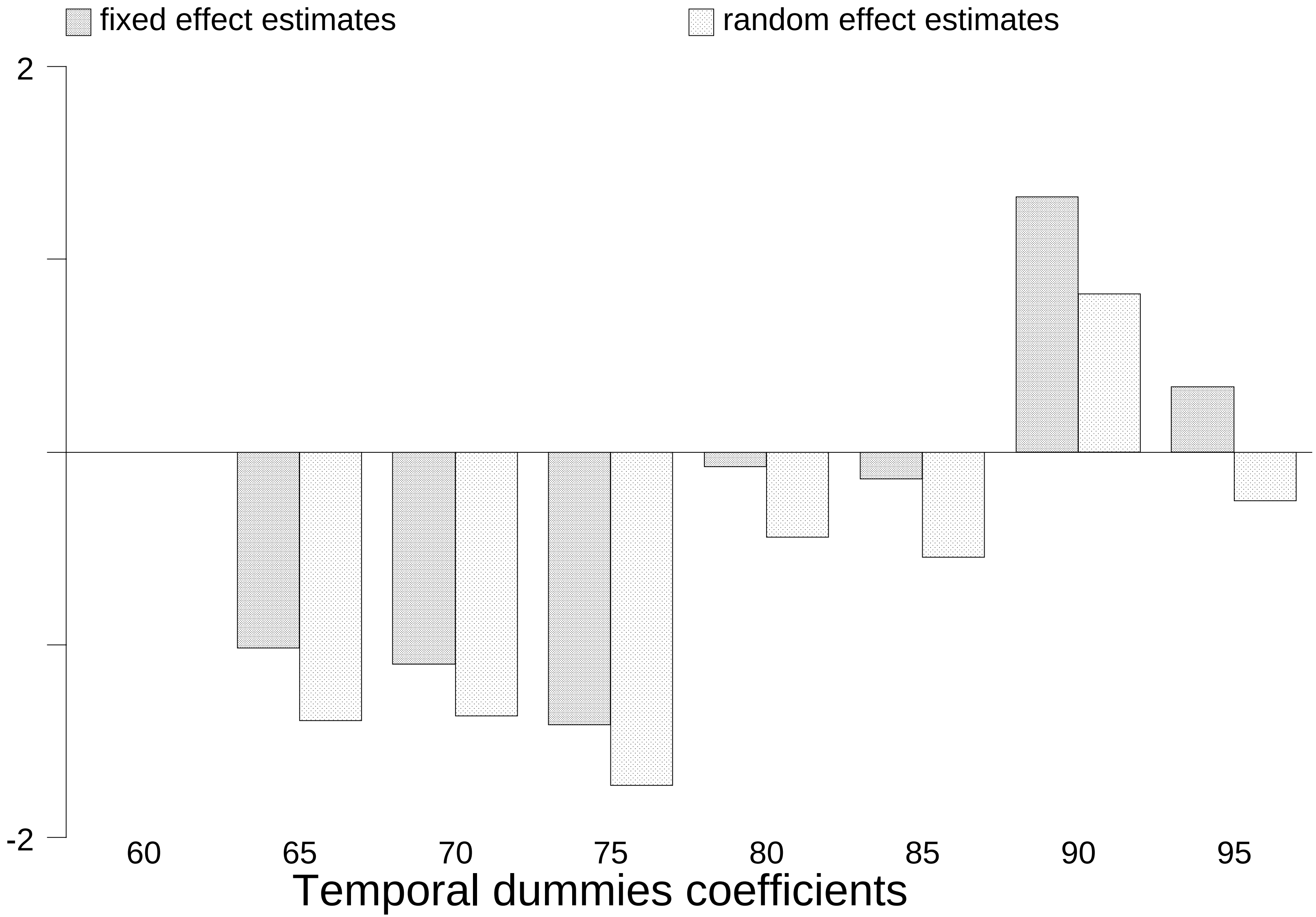
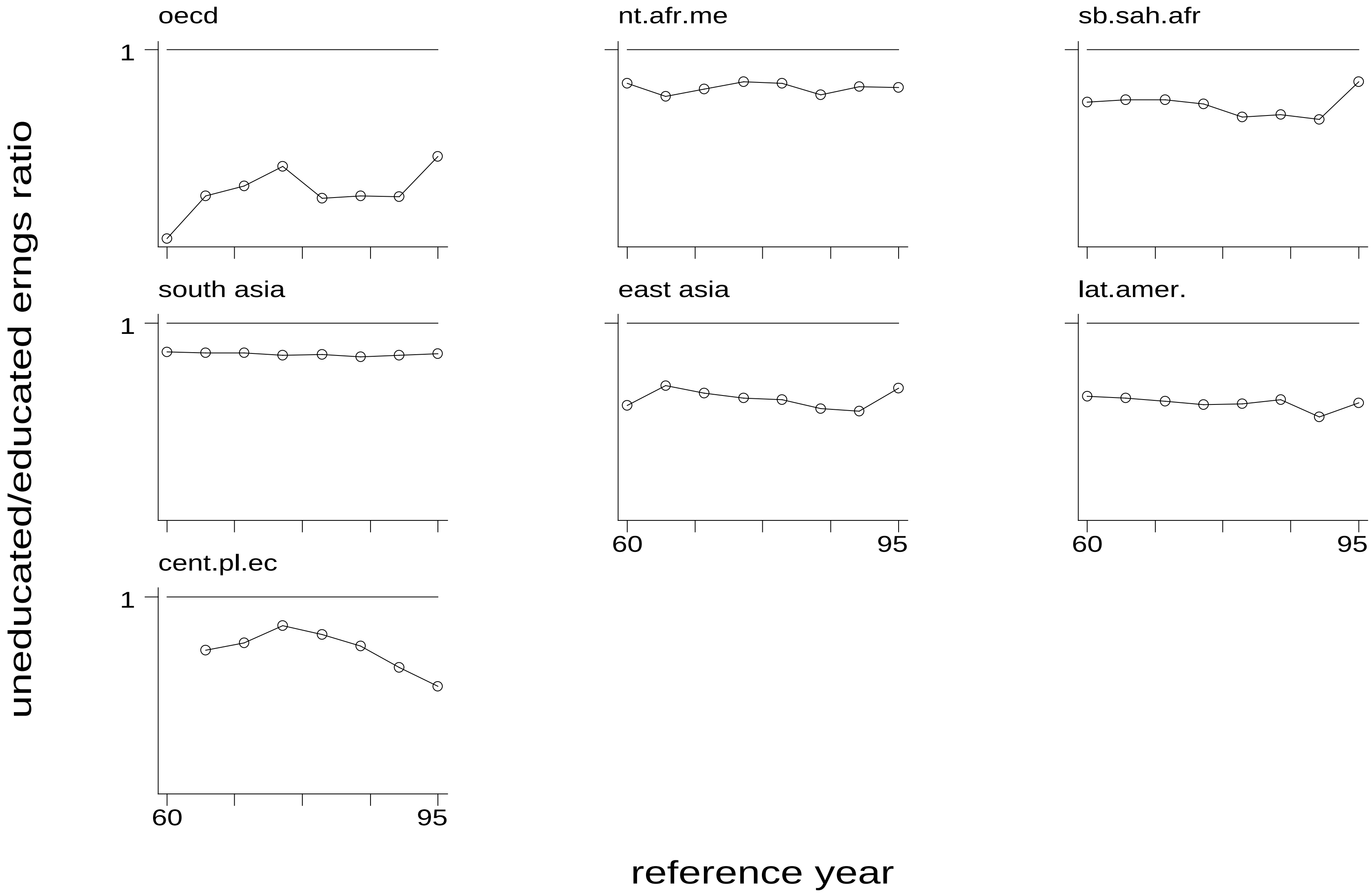
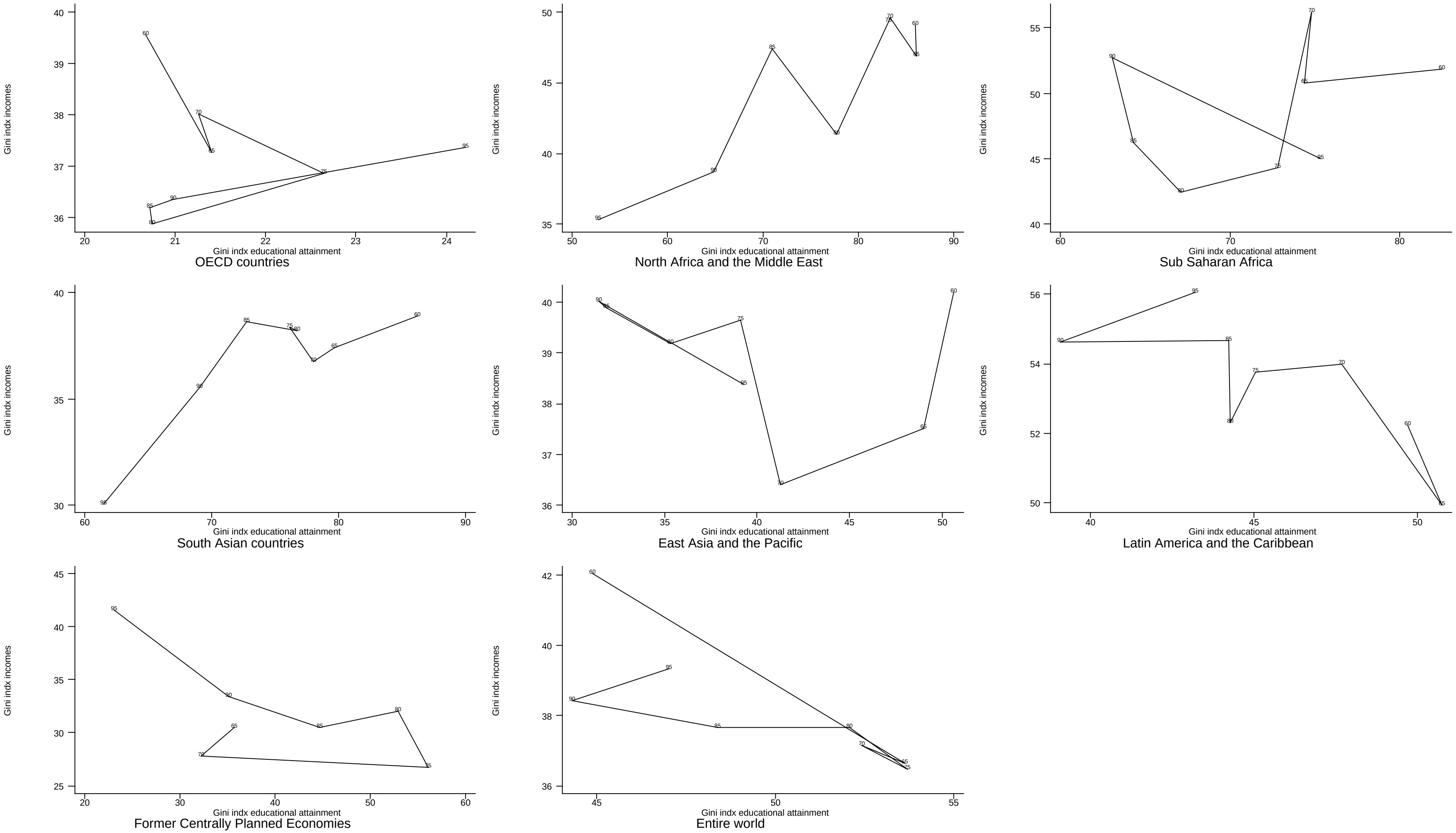

fixed effect estimates
random effect estimates
2
-2
60
65
70
75
80
85
90
95
Temporal dummies coefficients

Figure 5



Region/year averages (weights=population)

Figure 6



means - weights=population

Dynamics of inequality in incomes and education