Education, Growth and Income Inequality

Coen Teulings and Thijs van Rens*

First version: January 2001
This version: February 2003

Abstract

Estimates of the effect of education on GDP (the social return to education) have been hard to reconcile with micro evidence on the private return. We present a simple explanation that combines two ideas: imperfect substitution between worker types and endogenous skill biased technological progress. When types of workers are imperfect substitutes, the supply of human capital is negatively related to its return, and a higher education level compresses wage differentials. We use cross-country panel data on income inequality to estimate the private return and GDP data to estimate the social return. The results show that the private return falls by 1.5 percentage points when the average education level increases by a year, which is consistent with Katz and Murphy's [1992] estimate of the elasticity of substitution between worker types. We find no evidence for dynamics in the private return, and certainly not for a reversal of the negative effect as described in Acemoglu [2002]. The short run social return equals the private return, but the long run return is two times higher, providing evidence in favor of endogenous technological progress. The rise in education is the major cause of productivity growth over the sample period 1960-1990.

JEL classification: E20, J24, O10, O15

Under revision for resubmission to the *Review of Economics and Statistics.*

*We thank Mikael Lindahl and Daniele Checchi for sharing their datasets. We are also grateful to Robert Shimer, Giorgio Primiceri, Alan Krueger, Roland Bénabou, Mikael Lindahl, Miquel Pellicer, Maurice Bun, Miguel Portela, Richard Blundell, Steve Bond and participants of the 2002 annual meeting of the Society of Labor Economists for helpful comments and discussions. *Coen Teulings, Tinbergen Institute and Erasmus University Rotterdam, teulings@few.eur.nl. Thijs van Rens, Princeton University, tvanrens@princeton.edu.* Updates and data: http://www.princeton.edu/~tvanrens/educ.

1 Introduction

In a perfectly competitive world, the effect of an increase in the average education level of a country's workforce on log GDP, the social rate of return to education, equals the Mincerian private rate of return as estimated from data on individual wages. However, this prediction seems to be contradicted by the data. In a growth regression, education seems to have little effect on the level of GDP, but a strong effect on the growth rate. We present a simple model to reconcile both bodies of evidence. It combines two ideas: imperfect substitution between worker types and endogenous skill biased technological progress. Cross-country panel data on GDP have been routinely used for the analysis of the social return to education. Our idea is to use panel data on income inequality as a proxy for the private return. The empirical evidence on the joint evolution of the social and the private return to education is consistent with the model's prediction and with previous evidence on the degree of substitution between worker types, see Katz and Murphy [1992].

If workers with various levels of education were perfect substitutes, relative wages would be independent of the distribution of human capital. However, studies on the substitution between worker types, have shown that this is not the case. Then, a simple economic argument establishes that the Mincerian rate of return should be negatively related to the average years of education in the workforce. Raising average years of education makes low-skilled workers scarcer, raising their wages, while at the same time increasing the supply of highly educated workers and reducing their wages. This mechanism reduces the return to human capital. If externalities in education can be ignored, as suggested by a number of recent studies [Heckman and Klenow 1997; Acemoglu and Angrist 1999], the social rate of return to education equals the private rate, and the aggregate relation between GDP and education is a simple reflection of a Mincerian earnings function. Then, imperfect substitution between worker types has joint implications for GDP and income dispersion: we expect a negative second order effect of education on GDP and a related first order effect on inequality.

However, research on the cross country relation between education and GDP has documented an effect of education on GDP *growth*, not on its level [Benhabib and Spiegel 1994; Barro and Sala-i-Martin 1999]. The effect of changes in education on changes in GDP is insignificant in their regressions. These results have cast doubt on the relevance of the Mincer equation for the aggregate level, and have increased the popularity of human capital based endogenous growth models. In Barro and Sala-i-Martin for instance, a higher education level makes the labor force more able to deal with technological innovations, yielding a relation between the level of human capital and the growth of output.

To model imperfect subsitution between workers of different education levels, we use an assignment model with heterogeneous workers and heterogeneous jobs developed in Teulings [1995, 2002]. In this model, highly educated workers have a comparative advantage in complex jobs. The return to education is therefore higher in more complex jobs. When the supply of highly educated workers increases, there are insufficient complex jobs for them. High skilled workers have to do less complex jobs, where their human capital has a lower return. This yields a negative relation between the aggregate supply of education and its rate of return. We then extend this static Walrasian model, to allow for dynamics caused by firm's decisions to invest in new technologies. As in Acemoglu [2002], we argue that investments in new, skill biased technologies is more profitable when educated workers are more abundantly available. Since an increase in the average education level reduces the rate of return to education, it makes the application of more skill biased technologies profitable, and induces firms to invest in new technologies. We further assume that investing in knowledge is more human capital intensive than goods production. An increase in the average education level of the workforce will initially induce higher investments in new knowledge, but as this new knowledge enlarges the skill bias in the applied technology, the demand for human capital starts moving up, eroding the profitability of further investments in knowledge. This mechanism will cause the long-run social return to education to exceed the private return.

The evidence in Benhabib and Spiegel [1994] and Barro and Sala-i-Martin [1999] suggests an incredibly strong effect of education on GDP growth. Following Krueger and Lindahl [2000], we argue that their conclusions are at least partly due to measurement error, which attenuates the coefficient on the growth in education. However, Krueger and Lindahl's argument is not the whole story. Although they do not discuss this issue explicitly, the long run rate of return to education implied by their estimates is six times higher than the short run rate. We argue that two factors are crucial in understanding this finding. First, several studies have shown the importance of fixed country characteristics for GDP. Whether these effect are due to geography, as in Gallup, Sachs and Mellinger [1999], where proximity to the sea and a temperate climate are the driving forces, or to the better juridical institutions in countries with a more permanent involvement of European settlers, as in Acemoglu, Johnson, and Robinson [2001], they are likely to be correlated with education and will therefore bias the estimates in a pooled OLS regression. Countries with a favorable fixed effect are richer and can therefore invest more in human capital. Human capital variables pick up part of the favorable fixed effect, leading to overestimation of the long run effect of education. Second, the initial advantage in human capital increases over time due to skill biased technological progress. This gives the impression that education yields a higher growth of GDP, not a higher level. Since observed changes in education are perturbed by measurement error, skill biased technological change is hard to distinguish from endogenous growth. A combination of fixed effects, imperfect substitution and skill biased technological progress brings us much closer to a reconciliation of the GDP data and the Mincer equation.

Empirical research in this area is troubled by the issue of causality: does a higher education level lead to higher GDP or is it the other way around. Indeed, Bils and Klenow [1998] have argued that the posited causation from education to growth should be reversed. However, their arguments apply to the endogenous growth relation, and not to the Mincerian earnings function.[1] Our solution to the endogeneity problem relies

[1] Bils and Klenow argue that if endogenous growth is due to the role of education diffusing the most recent state of technology, then the education of new cohorts should be more valuable, leading to a

on time-lags in the causation from GDP to the average schooling level in the population. First the political system has to decide on spending of additional tax revenues on education. Then, new teachers have to be trained and schools have to be built. Only then the first cohorts can benefit from the improved training. It will then still take several years before these cohorts of better educated students enter the labor market and have a noticeable effect on the average level of education of the workforce. We argue therefore, that it is reasonable to assume that GDP affects education with a lag of at least 10 years, which is the time period we use in our regressions.

Our empirical results provide strong support for a negative relation between the supply of human capital and its return. Moreover, the estimation results for inequality and GDP are largely mutually consistent quantitatively. A one year increase in the stock of human capital reduces its return by 1.5 percentage points. This estimate is consistent with Katz and Murphy's [1992] estimate of the elasticity between low and high skilled workers in the US. We account for skill biased technological progress by entering cross effects of time dummies and education. Our estimates suggest skill biased technological change accounts for a 3% to 4% increase in the return to education per decade.

The estimates for the private return to education from inequality data are consistent with a sample of Mincerian returns to schooling estimated from microdata in several countries. We also find that in the short run, the social return to education approximately equals the private return, once imperfect substitution, skill biased technological progress and country-specific fixed effects are taken into account. This is a considerable advance from the growth literature, which has typically found that the effect of increases in education on growth is insignificant. We also find strong evidence that skill biased technological progress is endogenous: the long run social return to education is about two times higher than the short run return. However, we find no evidence for dynamics in the private rate of return beyond the initial drop after an increase in the average education level of the workforce. This contradicts the prediction in Acemoglu [2002] that an increase in education might raise the private return in the long run, because of increased incentives for investment in the invention of new skill biased technologies.

Despite the fact that our estimate of the long run social return is lower than in previous studies, it is still large, in fact larger than the actual GDP growth over the sample period. This relates our analysis to O'Neill [1995]. He asks the question why the huge investments in human capital by LDCs have not contributed to a convergence in GDP between LDCs and the industrialized world. His explanation relies on skill biased technological progress: "The recent shift in production techniques toward high-skilled labor has resulted in a substantial increase in the returns to education. This trend, when combined with the large disparities that still exist in education levels between the developed and less developed countries, has led to an increase in inequality despite the significant reduction in the education gap that has occurred over the last 20 years." [p.1299]. Our results confirm his analysis. Skill biased technological progress has shifted the terms of trade against developing countries, which produce commodities with a low capital intensity. Those countries that did not invest in human capital experienced

negative correlation between growth and the return to experience.

negative productivity growth.

The paper is structured as follows. In section 2, we present a simple Walrasian model with imperfect substitution between types of labor and endogenous technological progress. Section 3 discusses the data and presents the estimation results. Section 4 concludes.

2 Theoretical framework

2.1 The basic model

2.1.1 The structure of the economy

Consider the long run growth path of an economy with physical and human capital, along the lines of Teulings [1995, 2002]. Workers differ by their education level s, and tasks in the production process differ by their level of complexity c. Both s and c vary continuously along the real domain, so that we have an infinite number of types on both sides of the market. The supply of skill types s is exogenous in this model. It is assumed to be normally distributed: $s \sim N(S_t, V)$. We analyze the impact of changes in the average educational attainment of the workforce S_t on the economy. The variance of the skill distribution is assumed to be constant over time.

Each s-type worker can perform any c-type task. However, better educated workers have an *absolute advantage*: they are more productive in any task. In addition, they have a *comparative advantage* in more complex tasks. All markets are perfectly competitive. We can think of this economy as having two classes of firms: producers and composers. A producing firm produces a single c-type intermediate commodity associated with that c-type task. It sells its output on the market for intermediate commodities at a market price $P_t(c)$ at time t. A composing firm buys c-type commodities on the commodity markets and bundles them in a composite consumption (or investment) good by a Leontief technology. The c-type commodities are therefore demanded in fixed proportions.[2]

Production in a producing firm of type c is governed by a constant returns to scale Cobb Douglas production function with human and physical capital.

$$Y_t(c) = K_t(c)^{\alpha} H(c,s)^{1-\alpha} \tag{1}$$

where $Y_t(c)$ is production per worker of the intermediate commodities of type c, and $K_t(c)$ is the capital stock per worker. $H(c,s)$ is the productivity of workers with education level s in a c-type task. It satisfies:

$$\log H(c,s) = h(s-c)$$

[2] The distinction between two types of firms is the easiest way to present the model. Alternatively, one can think of the production process for the consumption good in a single firm (internalizing the markets for intermediate products) or having that all c-type commodities enter directly into the utility function. These interpretations yield exactly the same results.

with $h' > 0$ and $h'' \leq 0$. The restriction $h' > 0$ implies absolute advantage: an increase in s raises productivity in jobs of all complexity levels. The restriction $h'' \leq 0$ implies that the cross derivative of $\log H(c,s)$ is positive, yielding comparative advantage of highly educated workers in complex tasks: the relative productivity gain of an additional unit of s is increasing in c.

Firms choose the education level of their workers and the level of capital per worker as to maximize profits.

$$P_t(c) Y_t(c) - W_t(s) - RK_t(c) \tag{2}$$

where $W_t(s)$ is the market wage of a s-type worker and R is the rental rate of capital, which we assume to be constant over time. Since all markets are competitive, firms take wages and prices as given. Hence, the first order conditions of a c-type firm are given by

$$\begin{aligned} RK_t(c) &= \alpha P_t(c) Y_t(c) && (3) \\ W_t'(s_t(c)) &= (1-\alpha) P_t(c) Y_t(c) h'(s_t(c) - c) && (4) \end{aligned}$$

where $s_t(c)$ is the education level of the workforce in a c-type firm in market equilibrium.

The first order condition for capital (3) reflects the standard result for a Cobb-Douglas technology that the rental costs of capital are a fixed share α of output. Free entry of firms drives profits to zero, so that equation (2) combined with (3) implies that

$$W_t(s_t(c)) = (1-\alpha) P_t(c) Y_t(c) \tag{5}$$

Using this result, first order condition (4) becomes

$$w_t'(s_t(c)) = h'(s_t(c) - c) \tag{6}$$

where $w_t(s) \equiv \log W_t(s)$. It can be shown that $h'' \leq 0$ is sufficient for the second order conditions to be satisfied. Equation (6) has a simple interpretation. The left hand side is the Mincerian return to human capital, or from the point of view of the firm, the relative cost of the marginal unit of education of its workforce. The right hand side is the relative increase in labor productivity of the marginal unit of education. The first order condition states that in equilibrium both have to be equal.

Composing firms combine the c-type intermediate commodities by a Leontief technology into the composite consumption (or investment) commodity. Let Y_t denote the aggregate output of this composite commodity (or GDP) per worker. In a finite number of types world, a Leontief technology is characterized by a set of coefficients, one for each intermediate commodity type, indicating how many units of that type are required to produce one unit of output. In this infinite type world, the coefficients of the Leontief technology can be represented by a density function divided by an efficiency parameter F_t. A rise in F_t represents skill neutral technological progress: the same level of input $Y_t(c)$ yields more output Y_t. The ratio of the efficiency parameter and the density function of type c indicates how many units of that type are needed for the production of one unit of output. The distribution of input of intermediate commodities of type c

required for the production of Y_t is assumed to be normal $c \sim N(C_t, V)$. Hence, the input of type c at time t, denoted $Y_t(c)$, is given by:

$$Y_t(c) = \phi\left(\frac{c - C_t}{\sqrt{V}}\right)\frac{Y_t}{F_t} \tag{7}$$

where $\phi(.)$ denotes the standard normal density function. The mean of c, C_t, measures the average complexity level of the production process. A joint increase in both C_t and F_t is equivalent to skill biased technological progress: the demand for complex commodities, in the production of which highly educated workers have a comparative advantage, goes up relative to the demand for less complex products. The assumption that the variance of c equals the variance of the skill distribution V is obviously restrictive, but simplifies the subsequent analysis greatly. We will return to the implications of this assumption below (see the discussion after equation 14).

Equilibrium on commodity and labor markets in this economy is characterized by a set of wages $W_t(s)$ and prices $P_t(c)$ and an assignment rule of worker types to tasks $s_t(c)$, that satisfy zero profit condition (5), first order condition (6), and market clearing on the market for tasks of each complexity level. Market clearing requires that the demand for each c-type task equals its supply. Demand is given by expression (7). Supply equals the supply of workers of type $s_t(c)$ who produce that c-type, multiplied by their productivity $H(c, s_t(c))$ in producing that commodity. Substituting the normal density functions for the distributions of s and c and taking logaritms, the market clearing condition can be written as

$$y_t - \log F_t - \tfrac{1}{2}\frac{(c - C_t)^2}{V} = h(s_t(c) - c) - \tfrac{1}{2}\frac{(s_t(c) - S_t)^2}{V} + \log s_t'(c) \tag{8}$$

where $y_t = \log Y_t$. The final term on the right hand side, $\log s_t'(c)$, is the log of the Jacobian $ds_t(c)/dc = s_t'(c)$ for the transfer from a density function in skill levels $s_t(c)$ on the right hand side to a density function in complexity levels c on the left hand side. The term $-\frac{1}{2}\log V$ of the log normal density cancels on both sides because the variances are equal.

For this special case where the variances of the education and the complexity distributions are equal, differential equation (8) has an analytical solution:[3]

$$s_t(c) = c - C_t + S_t \tag{9}$$

Two observations are in place here. First, better skilled workers are assigned to more complex tasks, $s_t'(c) > 0$. This is what one would expect since they have a comparative advantage in these tasks. Second, holding c constant, the education level of a worker doing a c-type task rises when the mean education level of the workforce S_t goes up, and falls when the average complexity of the production process C_t rises.

[3] The initial condition that yields a unique solution to differential equation (8) is given by a transversality condition: for any other solution $\lim_{c\to\infty} s_t(c) = \overline{s} < \infty$ (implying that worker types $s > \overline{s}$ are not employed), $\lim_{c\uparrow\overline{c}} s_t(c) = \infty$ (so that there are no workers left to do tasks $c > \overline{c}$), $\lim_{c\to-\infty} s_t(c) = \underline{s}$, or $\lim_{c\downarrow\underline{c}} = -\infty$, all of which violate market clearing.

2.1.2 The firm's choice of technology

Suppose now that composing firms can choose the parameters of the Leontief technology that they use for the production of output from a set of technologies that is available at that particular point in time. More specifically, composing firms can choose the average complexity level of inputs C_t. The higher C_t, the higher of efficiency of the production process, but also the higher is the required input of complex tasks relative to simple tasks. This idea can be captured by letting the efficiency parameter F_t depend on C_t as well as the level of technological development which progresses exogenously over time:

$$\log F_t = f\left(C_t, t\right) \tag{10}$$

with $f_1 > 0$, $f_2 > 0$, $f_{11} < 0$, $f_{22} < 0$ and $f_{12} > 0$. As before, the effect of t on F_t can simply be interpreted as skill neutral technical progress. As time proceeds, the production process of composing firms becomes more efficient ($f_2 > 0$ excludes the possibility of technological regress). As noted before, an increase in C_t, which also increases F_t, represents skill biased technological progress. Because the cross-derivative f_{12} is positive, the reward for using more complex technologies increases over time.

Substituting (9) into differential equation (8) yields an expression for log output per capita.

$$y_t = h\left(S_t - C_t\right) + f\left(C_t, t\right) \tag{11}$$

Composing firms choose the average complexity level of their production process to maximize output given the supply of labor that is available. In this perfectly competitive world without externalities, the optimal technology maximizes y_t. The relevant first order condition reads:

$$-h'\left(S_t - C_t\right) + f_1\left(C_t, t\right) = 0 \tag{12}$$

Let $C_t = C\left(S_t, t\right)$ be the level of technology that satisfies this condition. The partial derivatives of this function can be calculated by implicit differentiation (leaving out arguments of all functions for convenience):

$$\begin{aligned} C_1 &= \frac{h''}{h'' + f_{11}} \in (0,1) \\ C_2 &= -\frac{f_{12}}{h'' + f_{11}} \in (0,\infty) \end{aligned} \tag{13}$$

The optimal complexity level of the production process C_t is increasing in the average education level of the workforce S_t. The mechanism underlying this relation is a general equilibrium effect: the higher the average level of education, the lower the return to human capital and the cheaper is the production of education intensive high c-tasks, thus making it attractive for firms to use higher C_t type of technologies. The positive relation between S_t and C_t corresponds Caselli and Coleman's [2002] finding of a positive correlation between computer use and the average education level of a country's workforce.

The optimal level of C_t increases over time, because $f_{12} > 0$. This captures the effect of skill biased technological progress. In the course of time, the reward for using more complex technologies increases as new and more efficient technologies become available. Hence, even at a constant average education level, composing firms will choose higher values of C_t as time goes by. The more complex production process raises the demand for skilled workers, causing the return to education to increase. This upward pressure can only be offset by a compensating increase in the supply of human capital S_t. This is Tinbergen's race between education and technology. If the set of new, more complex technologies grows faster than the supply of education, the return to education goes up, and vice versa.

2.1.3 Diminishing returns to education

What are the implications of this model for the return to education? Substitution of the equilibrium assignment rule (9) into equation (6) yields an expression for the evolution of the Mincerian return to human capital

$$w_t'(s) = w_t' = h'(S_t - C_t) \tag{14}$$

Equation (14) shows that the private rate of return to education, w_t' does not depend on s. Hence, $w_t(s)$ is linear in s. The implication that log wages depend linearly on years of schooling is consistent with a large body of evidence from the labor literature (see Card [1999] for an overview). In the context of the current model, the result depends crucially on the assumption that the variances of the education and complexity distributions are equal. Because of this feature $s_t'(c) = 1$ for all S_t and C_t so that the term $\log s_t'(c)$ drops out of equation (8) and $h(s_t(c) - c)$ is a constant independent of c. One can show that when the variance of the skill distribution is greater then the variance of the distribution of complexity levels, the Mincer equation is concave, and in the opposite case it is convex [Teulings 2002]. There is a simple intuition for this result. A larger variance of the skill distribution makes skill types skill types around the median more scarce compared to types in the tails of distribution. Hence, wages in the middle go up and wages in the tails go down. The assumption that the Mincer equation is linear is important for the interpretation of our empirical results, see the footnote on page 10.

The social rate of return to education is obtained by taking the derivative of equation (11) with respect to the average level of schooling

$$\frac{dy_t}{dS_t} \equiv y_t' = h'(S_t - C(S_t, t)) = w_t' \tag{15}$$

The indirect effect of S_t on output via $C(S_t, t)$ can be ignored by the envelope theorem, see equation (12). The social return to education equals the private return. This is what one would expect in a Walrasian world where all markets are perfectly competitive.

The private and social return to education vary with the average education level of the workforce. First, consider the case where the average complexity level C_t is exogenous. Then, taking a derivative of expression (15) holding C_t fixed, immediately shows that the assumption that $h'' \leq 0$ implies that the returns to education are diminishing. This result is due to the imperfect substitution between various s-types. The more negative h'', the less substitutable are workers with different education levels, and the stronger is the general equilibrium effect on the return to education.[4]

When firms choose the complexity level of the production process optimally, the story is slightly more complicated. Taking the total derivative of expression (15) we get

$$\frac{dy_t'}{dS_t} = \frac{dw_t'}{dS_t} = (1 - C_1)\, h'' = \frac{f_{11} h''}{h'' + f_{11}} < 0$$

Since both h'' and f_{11} are negative, the returns to schooling are dimishing. If either h'' or f_{11} is equal to zero, relative wages do not depend on the relative supplies of s-types and worker types are perfect substitutes. This perfect substitution can have two causes. If $h'' = 0$, the substitution occurs by shifts in the equilibrium assignment of workers to jobs $s_t(c)$: a shock to S_t shifts $s_t(c)$, see equation (9), but not $h'(S_t - C_t)$, since $h'' = 0$. If $f_{11} = 0$, then $\partial C_t / \partial S_t = 1$ by (13). In this case, substitution between different workers occurs by shifting the choice of technology. A shock to S_t is offset completely by a shift in the optimal technology $C(S_t, t)$, and hence $h'(S_t - C_t)$ remains unaffected.

A summary statistic for the degree of substitution between worker types is the *compression elasticity* γ. It is defined as the percentage reduction in the return to human capital per percent increase in the value of its stock:

$$\gamma \equiv -\frac{dw_t'/w_t'}{w_t' dS_t} = -(1 - C_1)\frac{h''}{h'^2} \tag{16}$$

The numerator in the first expression is the relative change in the return to education, the denominator is the relative change in the value of the stock of human capital, evaluated at its current rate of return w_t'. For the standard case where the are only two types of workers, the compression elasticity relates to the elasticity of substitution between high and low-skilled labor $\eta_{\text{low-high}}$ by the following relation, see Teulings [2002]

$$\gamma = \frac{1}{\eta_{\text{low-high}} D_t} \tag{17}$$

where D_t denotes the variance of log wages. Using Katz and Murphy's [1992] estimate of $\eta_{\text{low-high}} = 1.4$ and using a typical value for wage dispersion in the United States of $D_t \cong 0.36$, the compression elasticity is of the order of magnitude of 2 for the United States. We will use equations (16) and (17) to compare our estimates for the degree of substitution between worker types with Katz and Murphy's results.

[4] The testable implication of diminishing returns to education is that the second order effect of education on GDP is negative. However, this interpretation of the second order effect relies on the linearity of the Mincer equation (14) in s. If the Mincer equation were concave, the second order effect of S_t on log GDP would be negative even if worker types were perfect substitutes. Then, the declining return is due to a movement along the curve $w_t(s)$, instead of a movement of the curve itself. However, since there is abundant evidence on the linearity of the Mincerian earnings function from microdata, we shall interpret our estimation results under this assumption.

In summary, the basic model yields the following testable implications: the private and the social return to education are equal, and both are a negatively related to the average level of education in the economy, due to imperfect substitution between worker types. Hence, the size of this negative effect can be used as an estimate of degree of substitution between worker types.

2.2 The extended model with endogenous technology

2.2.1 The production of knowledge

Although the choice of the technology that is actually deployed by composing firms is endogenous in the basic model, the set of available technologies is exogenous. In this section, we consider an extension to the basic model, where the invention of new technologies requires effort. The set of available technologies depends on the stock of knowledge X_t available at time t. Hence, we replace the argument t by X_t in equation (10). Like in the basic model, composing firms produce the composite commodity. However, in the extended model these firms also engage in knowledge production. Let x_t denote the share of the workforce engaged in knowledge production. The remaining fraction $1-x_t$ of workers produce the consumption good. Following Uzawa [1965], Lucas [1988] and Barro and Sala-i-Martin [1999, chapter 5], we assume that the production of knowledge is more human capital intensive than the production of the consumption good. The parameter ε measures the excess human capital intensity of knowledge production: if C_t is the average complexity of tasks in the production of the consumption good, then the average complexity of tasks in knowledge production is $C_t + \varepsilon$. Since a share x_t of the workforce is engaged in knowledge production, the average complexity level of production in the economy equals $(1 - x_t)\, C_t + x_t\, (C_t + \varepsilon) = C_t + \varepsilon x_t$.[5] Hence, the production of the consumption good as in equation (11), can now be written as:

$$Y_t = (1 - x_t) \exp\left(h\left(S_t - C_t - \varepsilon x_t\right) + f\left(C_t, X_t\right)\right) \tag{18}$$

When firms invest more in new technologies (x_t goes up), current production decreases for two reasons: less workers are available for goods production, as measured by the factor $1 - x_t$, and the average productivity of workers in goods production goes down because knowledge production is more human capital intensive, measured by the term $-\varepsilon x_t$. As before, the complexity level of the production process is optimally chosen to maximize output in each period. Because the most highly educated workers are assigned to knowledge production, firms deploy less complex technologies during the investment phase: $C_t = C\left(S_t - \varepsilon x_t, X_t\right)$.

We allow for externalities in the investment in new knowledge. For that purpose, we have to distinguish between a specific composing firm j and all other composing firms.

[5] A structural way to model this is to consider an economy with two sectors: one for the production of the consumption good and one for knowledge production. Then, an arbitrage condition for each s-type determines the allocation of workers of that type over the two sectors. However, this would lead to substantial mathematical complexity. The model in the text can be interpreted either as a reduced form representation valid for small ε, or as a model of joint production of knowledge and the consumption good.

The evolution of the stock of knowledge available to firm j, X_{jt}, is determined partly by its own investment in new knowledge x_{jt} and partly by the average investment by all other firms $\vec{x}_{jt}$:

$$\dot{X}_{jt} = \theta x_{jt} + (1-\theta)\,\vec{x}_{jt} \tag{19}$$

where $\theta \in (0,1]$ measures the amount of knowledge spillovers.[6] If $\theta = 1$, firm j captures all the revenues from its own investment in new knowledge. For any value of $0 < \theta < 1$, firms capture only part of the revenues on their investment. Since all composing firms are assumed to be identical, in equilibrium they all invest the same amount in knowledge production: $\vec{x}_{jt} = x_{jt}$. Firms maximize the net present value of expected profits, taking the choices of all others firms as given. We assume firms' demand functions are iso-elastic. Under that assumption and with constant marginal costs, profits are a fixed share of value added. Profit maximization is therefore equivalent to the maximization of output. Firm j's intertemporal problem is to maximize the net present value of output over its investments in knowledge, subject to dynamic constraint (19).

$$\max_{x_{jt},X_{jt}} \int_{t=0}^{\infty} e^{-\rho t} Y_{jt} dt = \int_{t=0}^{\infty} e^{-\rho t}\,(1-x_{jt}) \exp\left(h\left(S_{jt} - C_{jt} - \varepsilon x_{jt}\right) + f\left(C_{jt}, X_{jt}\right)\right) dt$$

where ρ is the discount rate. Substituting $\vec{x}_{jt} = x_{jt}$ and dropping the j subscript since all firms are identical so that aggregate and firm level variables are the same, the first order conditions can be written as

$$\begin{aligned} \dot{X}_t &= x_t \\ (1-x_t)\,\theta\lambda_t &= P_{xt} Y_t \quad \text{where } P_{xt} \equiv 1 + (1-x_t)\,\varepsilon h' \\ \dot{\lambda}_t &= \rho\lambda_t - f_2 Y_t \end{aligned} \tag{20}$$

where λ_t is the Lagrange multiplier on the law of motion for X_t. Indirect effects of x_t and X_t via C_t drop out of the first order conditions by the envelope theorem. P_{xt} is the relative price of investments in new technology. It is greater than unity since production of knowledge is more human capital intensive than production of the consumption good, and therefore depends positively on the return to human capital h'.

The total derivative of the second equation yields an expression for $\dot{\lambda}_t$. Substituting λ_t and $\dot{\lambda}_t$ as well as $x_t = \dot{X}_t$ into the third equation gives a second order (non-linear) differential equation in the level of technological development:

$$P_{xxt}\dot{x}_t = \rho P_{xt} - \theta f_2 - \left[(P_{xt} - \theta)\,f_2 - (1-x_t)\,C_2 \varepsilon h''\right]\dot{X}_t \tag{21}$$

where P_{xxt} can be interpreted roughly as the effect of the level of investment in knowledge x_t on its price P_{xt}.

$$P_{xxt} \equiv -\varepsilon h'\,(1 + P_{xt}) - (1-x_t)\,(1-C_1)\,\varepsilon^2 h''$$

[6] We could generalize this specification by allowing $\dot{X}_{jt}$ to depend both on investments in knowledge and independently on time. We will allow for this more general specification in our empirical application but refrain from it here for simplicity.

If $P_{xxt} < 0$, the marginal cost of investment is falling. In that case, the economy would immediately jump to its new equilibrium stock of knowledge by a massive instantaneous investment. This is clearly not realistic scenario. In the context of the model presented here, the human capital intensity of knowledge production leads to increasing marginal cost: the higher x_t, the greater the demand for human capital, and hence the higher the cost of further investment, P_{xt}. This effect is captured by the second term in the expression for P_{xxt}: $-(1-x_t)(1-C_1)\varepsilon^2 h'' > 0$. For a sufficiently high value of h'', this effect dominates the first term, so that $P_{xxt} > 0$, which is what we assume in what follows.

2.2.2 The long run effects

The steady state level of technological development (given a certain education level of the workforce) is obtained by setting $\dot{x}_t = \dot{X}_t = 0$. Let a bar over a variable denote its long run steady state value. Then:

$$\rho \bar{P}_x - \theta \bar{f}_2 \equiv Q\left(S, \bar{X}\right) = 0 \tag{22}$$

This condition states that the cost of a unit of investment in knowledge, $\bar{P}_x$, must be equal to the annuity value of revenues from that investment for the investing firm, $\theta \bar{f}_2 / \rho$. The parameter θ accounts for that the investing firm receives only a share θ of the revenues. Implicitly differentiating equation (22) we obtain the following expressions for the partial derivatives of the function $Q\left(S, \bar{X}\right)$:

$$\begin{aligned} Q_1 &= \rho\varepsilon\bar{h}''\left(1-\bar{C}_1\right) - \theta\bar{f}_{12}\bar{C}_1 < 0 \\ Q_2 &= -\rho\varepsilon\bar{h}''\bar{C}_2 - \theta\left(\bar{f}_{12}\bar{C}_2 + \bar{f}_{22}\right) \end{aligned}$$

Existence of a long run equilibrium requires that the net cost of investing in knowledge is increasing in the stock of knowledge: $Q_2 > 0$. If this condition is not satisfied, an investment in the stock of knowledge raises productivity at an increasing rate. Because a higher level of technology raises the productivity of skilled labor, $f_{22} < 0$ by itself is not sufficient to guarantee diminishing returns to knowledge; f_{22} needs to be negative enough to outweigh the positive cross-effect $f_{12}C_2$ through technological progress on productivity. A higher stock of knowledge raises the return to education and therefore the cost of investing in more knowledge, which works in the right direction, as it makes the cost of investment increasing in the stock of knowledge. With falling cost of new knowledge, the resource constraint of the economy is no longer binding. Since that implication seems at odds with the world we observe, we impose the condition $Q_2 > 0$ in what follows.

Equation (22) establishes a steady state relation between S and $\bar{X}$, which we denote $\bar{X}(S)$. Implicitly differentiating equation (22) yields an expression for the derivative $\bar{X}'(S)$:

$$\bar{X}' = -\frac{Q_1}{Q_2} > 0 \tag{23}$$

The steady state stock of knowledge is an increasing function of the average level of education in the workforce. The reason is that technology is skill biased: the greater the stock of available human capital, the more profitable it is to invest in new technology. The effect of education on the long run social rate of return to education consists of both the direct effect of the larger average level of human capital and the indirect effect of the larger stock of knowledge. Differentiating equation (18), where again the effect via $\bar{C}$ drops out by the envelope theorem, we get:

$$\bar{y}' = \bar{h}' + \bar{f}_2 \bar{X}' = \bar{w}' + \frac{\rho}{\theta} \bar{P}_x \bar{X}' > \bar{w}' \tag{24}$$

where we use condition (22) to substitute for $\bar{f}_2$. Unlike in the basic model, the long run social return to education $\bar{y}'$ exceeds the private return $\bar{w}'$. The smaller the share θ of revenues from new inventions that accrues to the inventor, the larger this difference. However, even if there are no knowledge spillovers, $\theta = 1$, the social rate of return exceeds the private rate. This result simply reflects the return on investments in new knowledge. A once and for all shock to the average education level of the workforce of dS leads to a accumulated investment in new knowledge of $\bar{P}_x dX = \bar{P}_x \bar{X}' dS$, raising the long run level of output by $\rho \bar{X}' dS$ because the marginal product of more knowledge equals the market rate of return ρ. Notice that we touch upon an accounting issue here: our measure of GDP does not include investments in new knowledge, see equation (18). To the extent that these investments are non-tangible, this seems to be a realistic representation of real life accounting practices.

The effect of a unit shock in S on the long run private return to education is obtained by differentiating equation (14) and evaluating in the steady state:

$$\frac{d\bar{w}'}{dS} = \left(1 - \bar{C}_1 - \bar{C}_2 \bar{X}'\right) \bar{h}'' = -\left(1 - \frac{\bar{C}_2}{1 - \bar{C}_1} \bar{X}'\right) \bar{\gamma} \bar{w}'^2 > -\bar{\gamma} \bar{w}'^2 \tag{25}$$

Contrary to the basic model, the long run effect of an increase in the average level of education on the private return to education does not need to be negative. The counteracting mechanism is similar to that in Acemoglu [2002]: a greater supply of human capital induces investments in new technology. Since innovations to technology are skill biased, they raise the private return to human capital. These innovations can more than offset the effect of the initial increase in the supply of human capital. Since we have no theoretical prediction on the size of $\frac{\bar{C}_2}{1-\bar{C}_1} \bar{X}'$, the net long run effect of the average education level on the private return is an empirical question. Whether positive or negative, $-d\bar{w}'/dS$ is in any case smaller than $\bar{\gamma} \bar{w}'^2$, the decrease in the return to education due to imperfect substitution. Hence, in the extended model with endogenous technological development $-d\bar{w}'/dS$ is a lower bound for the compression elasticity since the compression of the wage distribution is partly offset by endogenous skill biased technological change.

Differentiating expression (24) gives the long run effect on the long run social return:

$$\frac{d\bar{y}'}{dS} = \frac{d\bar{w}'}{dS} + \bar{f}_2 \bar{X}'' \lesseqgtr \frac{d\bar{w}'}{dS} \tag{26}$$

The difference in the long run effects on the social and the private return depends on the sign of $\bar{X}''$, which in turn depends on the higher order derivatives of h and f on which we have no priors.

Summarizing, contrary to basic model, the long run social rate of return to education $\bar{y}'$ exceeds the private return $\bar{w}'$ because of investments in new technologies, even if there are no externalities in the production of knowledge. The sign of the long run effect of a shock in S on the social and the private return to education, $d\bar{y}'/dS$ and $d\bar{w}'/dS$, is ambiguous. However, the effect on the private rate is less negative in the basic model than in the extended model. In the basic model, the negative effect reflects the impact of the compression elasticity. In the extended model, there is an offsetting effect because a higher level of education induces additional investment in the stock of knowledge, leading to further skill biased technological progress.

2.2.3 Transition dynamics

Linearizing differential equation (21) around the steady state yields:

$$\bar{P}_{xxt}\dot{x}_t = Q_2 \left(X_t - \bar{X}\right) + \psi \dot{X}_t$$

where

$$\psi \equiv \bar{f}_2 \bar{C}_1 - \varepsilon^2 \left(1 - \bar{C}_1\right) \bar{h}'' - \left(\bar{P}_x - \theta\right) \bar{f}_2 + \varepsilon \bar{C}_2 \bar{h}'' \lesseqgtr 0$$

The parameter restrictions $P_{xxt} > 0$ and $Q_2 > 0$ are sufficient for this linear second order differential equation to a unique stable solution.[7] The approximate solution of the dynamics of the system around the steady state can be written as $X_t - \bar{X} = Ae^{-\lambda t}$, where A is an integration constant and where

$$\lambda = \frac{\psi + \sqrt{\psi^2 + 4\bar{P}_{xxt}Q_2}}{2\bar{P}_{xxt}} > 0$$

The level of technology converges to its long run steady state level according to an exponential process at a rate λ. All other variables follow the same exponential adjustment rules close to the steady state. We use these adjustment paths to calculate the immediate effect of a permanent shock in S_t on the endogeous variables of the model.

Consider a permanent increase of the average education level by one year. As discussed in the previous section, the long run effect of such a shock in S_t on the stock of knowledge is $\bar{X}'dS = \bar{X}'$. Since X_t can only adjust slowly there is no immediate effect on the level of technology. Plugging in $\bar{X} = X_0 + \bar{X}'$ we can write the solution as:

$$X_t - \bar{X} = -\bar{X}'e^{-\lambda t} \tag{27}$$

The derivative of X_t with respect to time yields the transition path of investment in new knowledge: $x_t = \lambda \bar{X}' e^{-\lambda t}$. Investment jumps up by $\lambda \bar{X}'$ immediately after the shock in S, and then gradually converges to its steady state value $x_t = 0$. Then, the immediate effect on the private return to education can be derived from equation (14):

$$w_0' - \bar{w}_0' = \left(1 - \varepsilon\lambda\bar{X}'\right)\left(1 - \bar{C}_1\right)\bar{h}'' = -\left(1 - \varepsilon\lambda\bar{X}'\right)\bar{\gamma}\bar{w}'^2 \tag{28}$$

[7] Both roots of the characteristic equation are real, one is always positive and the other negative.

where $\bar{w}'_0$ refers to the long run private return *before* the shock in S. Compared to the basic model, the negative immediate effect of the shock is partly offset by the jump in investments in knowledge. The human capital intensity of this investment drives up the private return. The net effect of both forces is ambiguous, depending on the sign of $1 - \varepsilon\lambda\bar{X}'$. The immediate effect effect on output can be calculated from expression (18)

$$y_0 - \bar{y}_0 = w'_0 - \lambda\bar{X}' \tag{29}$$

where $\bar{y}_0$ was the steady state level before the shock in S occurred. Like in the case of the private return, the investment in new knowledge reduces the immediate effect of the shock. Again the net effect of both forces is ambigous. However, an unambiguous implication of the model is that $y_0 - \bar{y}_0 < w'_0$: the short run social return is smaller than the private return because some workers are assigned to knowledge production. We shall test this implication.

The direction of changes in the private return *after* the initial jump is also ambiguous. On the one hand, the rate of investment declines when the stock of knowledge converges to its new steady state value, reducing the demand for human capital and therefore the private return. On the other hand, the greater stock of knowledge induces the deployment of more skill biased technologies, raising the demand for human capital and hence its return. Both effects might cancel, so that the there is only an initial effect on the private return and it remains constant afterwards. In any case, changes in the private rate of return following the initial response will be much smaller than changes in GDP, where the two effects, a greater stock of knowledge and a lower level of investment in new knowledge, work in the same direction.

Summarizing, the immediate effect of a shock in S_t on the private rate of return and on GDP is ambiguous. However, the effect on GDP is smaller than the private return. In the subsequent transition path, GDP rises unambiguously, while the private rate of return can either rise of fall, depending on whether the effect of either the greater stock of knowledge or the lower level of investment in knowledge dominates.

3 Empirical evidence

3.1 Direct evidence on the private return to education

Before turning to our main estimation results we present some direct evidence for imperfect substitution between worker types. The main hypothesis underlying this paper is that if workers of different education levels are imperfect substitutes, an increase in the supply of human capital will decrease its return. Linearizing equation (14) we get

$$w'_{jt} = \alpha_{1t} - \alpha_2 S_{jt} \tag{30}$$

where j indexes countries. Bils and Klenow [1998] collected a cross section sample of estimated private returns to education from microdata for a number of countries. These data that are listed in table I. We have plotted the return to education against the average schooling level in figure I. There is a clear negative relationship between the two variables. The simple regression estimates in table II confirm this. The crucial coefficient α_1 is always significant at the 1% level.

Based on these estimates, the private return to education is about 15% for countries with an education level of zero, and decreases by about 0.7% for every year of education. The private rate of return to education is 11% for the average education level of 5.3 years in 1990, while it is 7% for the US, with an average education level of 12 years. The parameter estimates for α_1 and α_2 can be used to calculate the compression elasticity as measure of the degree of imperfection in the substitution between types of labor, see equation (16):

$$\gamma_{jt} \equiv \frac{dw'_{jt}/dS_{jt}}{w'^2_{jt}} = \frac{\alpha_2}{(\alpha_{1t} - \alpha_2 S_{jt})^2} \tag{31}$$

For $\alpha_2 S_{jt} < \alpha_{1t}$, the compression elasticity is monotonically increasing in the average education level S_{jt}. The compression elasticity is equal to 0.60 for the average education level in 1990, while it is 3.44 for the United States. The latter number compares reasonably well to the value of 2 based on Katz and Murphy's [1992] estimate of the elasticity of substitution between high school and college graduates for US, see equation (17). The time dummies suggest the presence of skill biased technological progress from 1985 to 1990, raising the return to human capital by 4%. Weighing countries by log GDP per worker or log population size as in columns (3) and (5) does not affect these conclusions, nor are the results driven by outlier Jamaica, see columns (2), (4) and (6).

The estimation results in table II have a couple of drawbacks. First, they are based on cross section variation between countries only. Because we have only one estimate for the return to schooling in each country, it is impossible to control for country-specific effects. For example, if good institutions to protect property rights both favor the accumulation of human capital and reduce rent extraction, it is impossible to disentangle the two effects on the return to education from cross section data alone. Second, a cross section does not allow inference about the dynamics of the effect of education on its return. Hence, in the remainder of this section we focus on an analysis of wage dispersion and GDP, for which panel data are available.

3.2 Empirical specification for inequality and GDP

Since the accumulation of human capital reduces its return, it compresses the wage distribution. Figure II indeed documents a strong positive relation between inequality and the return to education. Because education clearly is not the only factor yielding wage differentials between workers, we extend the Mincer equation (6) to allow for other worker characteristics:

$$w_{jt}(s, u) = \alpha_{0jt} + w'_{jt}s + \sigma u \tag{32}$$

where u is a standard normal random variable representing other worker characteristics and σ is their standard deviation. Years of schooling s and other worker characteristics u are assumed to be jointly normally distributed, with correlation ρ (making the wage distribution log-normal). Under these assumptions, the variance of log wages or wage dispersion D_{jt} is given by:

$$D_{jt} = w_{jt}'^2 V + 2\rho\sigma w_{jt}' V^{1/2} + \sigma^2$$

where V is the variance of the education distribution. Using equation (32) to substitute for w_{jt}', we get the following expression for wage dispersion:

$$D_{jt} = \theta_{0t} - \theta_{1t} S_{jt} + \theta_2 S_{jt}^2 \tag{34}$$

where:

$$\begin{aligned}
\theta_{0t} &\equiv \alpha_{1t}^2 V + 2\alpha_{1t} V^{1/2} \sigma\rho + \sigma^2 \\
\theta_{1t} &\equiv 2\alpha_2 \alpha_{1t} V + 2\alpha_2 V^{1/2} \sigma\rho \\
\theta_2 &\equiv \alpha_2^2 V
\end{aligned}$$

Notice that both θ_{1t} and θ_2 would be zero if $\alpha_2 = 0$. Within the framework of the basic model, a proper test of perfect substitution between workers types is the joint restriction $\theta_{1t} = \theta_2 = 0$. The θ parameters depend on the variance of the education distribution V, which has been assumed to be constant over time and across countries. It is difficult to provide a structural model including the effect of variation in V_{jt}, since the Mincer equation is no longer linear in that case, see Teulings [2002]. We will adopt a pragmatic approach, by adding an additive control term $\theta_3 V_{jt}$ to equation (34). If we assume that capital income is distributed proportional to labor income, so that the log wage distribution and the log income distribution differ only by their first moment, then equation (34) holds for income inequality as well. Panel data on income inequality are available, so we can estimate this equation.

The theoretical framework has led to a number of predictions regarding the dynamics of the private and the social return to education in response to a shock to the average education level. Essentially, the model predicts that in response to shocks to the average schooling level in a country, both the private return to education and output converge to their new steady state levels according to an exponential adjustment rule. By the argument above, also inequality will follow such an exponential adjustment rule. These dynamic responses can be estimated from the following equations:[8]

$$\begin{aligned}
D_{jt} &= \theta_{0t} - \theta_{1t} S_{jt} + \theta_2 S_{jt}^2 + \bar{\theta}_{1t} S_{jt-1} - \bar{\theta}_2 S_{jt-1}^2 + \phi D_{jt-1} + u_{jt} && (35) \\
y_{jt} &= \beta_{0t} + \beta_{1t} S_{jt} - \tfrac{1}{2}\beta_2 S_{jt}^2 - \bar{\beta}_{1t} S_{jt-1} + \tfrac{1}{2}\bar{\beta}_2 S_{jt-1}^2 + \psi y_{jt-1} + v_{jt} && (36)
\end{aligned}$$

where u_{jt} and v_{jt} are error terms. We allow θ_{0t}, β_{0t} and β_{1t} to depend on time, the first two to account for skill neutral technological progress and the last one to account for skill biased technological progress. Equation (35) is the dynamic version of expression

[8] Notice that equation (36) for GDP can be rewritten in error correction form as

$$\Delta y_{jt} = \tilde{\beta}_{0t} + \beta_{1t} \Delta S_{jt} - \tfrac{1}{2}\beta_2 \Delta S_{jt}^2 + \left(\beta_{1t} - \bar{\beta}_{1t}\right) S_{jt-1} - \tfrac{1}{2}\left(\beta_2 - \bar{\beta}_2\right) S_{jt-1}^2 - (1-\psi)\, y_{jt-1} + v_{jt}$$

This specification is a growth regression, which has been estimated many times in the literature.

(34), and (36) follows from expression (18). The parameters in the equation for output are defined such that the expression for the short run social return to education

$$y'_{jt} = \beta_{1t} - \beta_2 S_{jt} \tag{37}$$

is easily comparable to expression (30) for the private return $w'_{jt} = \alpha_{1t} - \alpha_2 S_{jt}$.

The short run (first order) effect of an increase in S_{jt} on inequality and output is given by θ_{1t} and β_{1t} respectively, whereas the long run effects can be calculated as $\left(\theta_{1t} - \bar{\theta}_{1t}\right) / (1-\phi)$ and $\left(\beta_{1t} - \bar{\beta}_{1t}\right) / (1-\psi)$. If the private and social return to education are constant over time, the estimated long run coefficients will not necessarily be zero because there may be autocorrelation in the part of inequality and GDP that is not explained by education, but the long run effects will equal the short run effects.[9] This would be evidence in favor of the basic model where the set of available technologies is exogenous. If at least one of the restrictions $\theta_{1t} = \left(\theta_{1t} - \bar{\theta}_{1t}\right) / (1-\phi)$ and $\beta_{1t} = \left(\beta_{1t} - \bar{\beta}_{1t}\right) / (1-\psi)$ is rejected, this is evidence against the basic model and in favor of the extended model.

Finally, consider the role of physical capital in a regression like (36). The Cobb-Douglas production function (1) with equal shares of capital α for each c-type producing firm, generates the standard result that the rental costs of capital are a fixed share α of output as in first order condition (3). Aggregating over all firms, that result implies that log aggregate capital differs from log GDP only by a constant:

$$k_t = y_t + \log \alpha - r$$

Hence, we cannot the establish the contribution of human and physical capital to production by directly estimating the log of an aggregate version of the production function (1), since log physical capital k_t and log human capital $h\left(S_t - C_t\right)$ are perfectly collinear. Although this is probably an extreme case, which depends on the specific assumptions on the production function and the return to capital r being constant, it helps to understand an actual empirical problem. Given the collinearity of K and H and in the presence of measurement error in both variables, the relative magnitudes of the coefficients of both factor inputs merely reflect the precision of their measurement. Krueger and Lindahl [2000] argue that capital data are correlated to output by construction, since investment data figure in both series. Hence, measurement error in both series is likely to be correlated. This explains why they find the estimated α to be much higher than one would expect on the basis of conventional estimates of the capital share in output

[9] Consider for instance the inequality equation. Suppose that the true relation between education and inequality is static as in equation (34), so the model is given by

$$D_{jt} = \theta_{0t} - \theta_{1t} S_{jt} + \theta_2 S_{jt}^2 + u_{jt}$$

where the error term is first order autocorrelated $u_{jt} = \rho u_{jt-1} + \varepsilon_{jt}$. Then, we can rearrange to get an expression very much like equation (35)

$$D_{jt} = \tilde{\theta}_{0t} - \theta_{1t} S_{jt} + \theta_2 S_{jt}^2 + \rho\theta_{1t} S_{jt} - \rho\theta_2 S_{jt}^2 + \rho D_{jt} + \varepsilon_{jt}$$

where the parameter restrictions imply that the long run effects equal the short run effects.

of about 0.35. The coefficients from these regressions are therefore extremely difficult to interpret. We shall omit capital from all our regressions and apply the reduced form equation (36) in our empirical application.[10]

3.3 Identification and estimation

A problem in the estimation of equation (36) for GDP and (35) for wage dispersion is reverse causality: does a rise in education cause growth in GDP, or is it the other way around and does growth cause rising educational attainment? Our approach to this problem is to use the dynamic structure of the effects. We assume that contemporaneous changes in GDP do not affect the average education level of the workforce in that same observation period. As explained in the introduction, we follow Krueger and Lindahl [2000] by using a 10 year timeframe to alleviate the measurement error bias in the coefficient on ΔS_{jt}. To be able to interpret our estimates, we therefore need to assume that a change in GDP takes at least 10 years to affect the average education level in a country. This seems areasonable assumption. It takes time before an increase in GDP leads to an increase in the budget of the education system. Then, new schools have to be built or new teachers to be trained. Finally, the new generation of students that benefits from the increased expenditures on schooling takes time to finish school and enter the labor market.

Estimating (35) and (36) by OLS is consistent and efficient if the error terms are true innovations. However, the problem becomes more complicated if we want to allow for country-specific fixed effects. For clarity of exposition, focus on equation (36) for GDP, and suppose that

$$v_{jt} = f_j + \varepsilon_{jt}$$

where f_j is the fixed country effect and ε_{jt} is an innovation in GDP. The assumption on the time lag in the reverse effect of GDP on education implies:

$$E\left[S_{jt-s}\varepsilon_{jt}\right] = 0 \quad \text{for } s \geq 0$$

Clearly now OLS is inconsistent, because y_{jt} is correlated with f_j. First differencing equation eliminates the fixed effect, but now the component $\varepsilon_{j,t-1}$ of $\Delta\varepsilon_{jt}$ is correlated with $\Delta y_{j,t-1}$ (and possibly also with ΔS_{jt} via the reverse causality equation). A consistent and efficient GMM estimator uses the following moment conditions:

$$\begin{aligned} E\left[y_{jt-s}\Delta\varepsilon_{jt}\right] &= 0 \quad \text{for } s \geq 2 \\ E\left[S_{jt-s}\Delta\varepsilon_{jt}\right] &= 0 \quad \text{for } s \geq 1 \end{aligned}$$

[10] Alternatively, we could set the coefficient for capital at some fixed value, as Krueger and Lindahl [2000] do. Which procedure is most efficient depends on the type of deviations of the assumptions one thinks are most relevant. If there is measurement error in the capital data or if the capital share is constant over time, but varies between countries, then omitting capital as an explanatory variable is most efficient. If the long run return to capital varies over time, it is preferrable to set the contribution of capital to some fixed value. Both methods fail if the capital share varies both between and within countries. We tried both approaches and found little difference in the results.

These moment condition give rise to the dynamic panel data estimator set out in Arellano and Bond [1991].

Since we use 10 year time intervals, the time dimension of our panel is very short (at most four periods). As shown by Blundell and Bond [1998], we can realize a substantial efficiency gain if we are prepared to make the additional assumption that the country-specific fixed effect in GDP is uncorrelated with innovations in the education level:

$$E\left[f_j \Delta S_{jt}\right] = 0$$

Notice that the assumption is much weaker than $E\left[f_j S_{jt}\right] = 0$. It allows the fixed effect in GDP to affect the *level* of education, but not on its growth rate. Under this assumption, two additional sets of moment conditions are available:

$$\begin{aligned} E\left[v_{jt} \Delta y_{jt-s}\right] &= 0 \quad \text{for } s \geq 1 \\ E\left[v_{jt} \Delta S_{jt-s}\right] &= 0 \quad \text{for } s \geq 0 \end{aligned}$$

These additional moments conditions give rise to the system estimator proposed by Blundell and Bond. Both the Arellano-Bond and the Blundell-Bond system estimator are implemented using the DPD package for Ox [Doornik et.al. 2002].

3.4 Data sources

Our empirical analysis is largely based on data from two sources: the Barro and Lee [1996, 1993] data on educational attainment and the Deininger and Squire [1996] data on income inequality. These datasets were supplemented with data on real GDP per worker from the Penn World Table [Summers and Heston 1991] mark 5.6a.

The Barro and Lee dataset contains detailed data on educational attainment for 114 countries for the period 1960-1990 in intervals of 5 years. Barro and Lee report the fraction of the population that attained a certain education level, as well as the average duration of this education level. They use these data to construct the average education level of the population in years. We also calculate a rough estimate of the variance of the education distribution.[11]

[11] Barro and Lee calculate average years of education from attainment data (percentage of the population that have attained a certain level of schooling) combined with data on the typical duration of each level of schooling [1996, p.218]. We can express the calculation as:

$$S = f_{\text{prim}} S_{\text{prim}} + f_{\text{sec}} \left(D_{\text{prim}} + S_{\text{sec}}\right) + f_{\text{high}} \left(D_{\text{prim}} + D_{\text{sec}} + S_{\text{high}}\right)$$

where S is average years of schooling in the total population, f_{level} is the fraction of the population that has attained a certain education level (no education, primary education, secondary education or higher education), D_{level} is the typical duration of the different education levels, and S_{level} is the average duration of a certain education level for those people that have not continued to attain a higher education level. Intuitively $S_{\text{level}} < D_{\text{level}}$ due to early drop-out. The calculation of average years of schooling in this expression is just an expected value, which suggests the following proxy for the variance in education within each country, cf. Checchi [1999]:

$$V(S) = f_{\text{prim}} S_{\text{prim}}^2 + f_{\text{sec}} \left(D_{\text{prim}} + S_{\text{sec}}\right)^2 + f_{\text{high}} \left(D_{\text{prim}} + D_{\text{sec}} + S_{\text{high}}\right)^2 - S^2$$

Deininger and Squire [1996] use results from a large number of studies and assess their comparability. Their dataset contains Gini coefficients of the income distribution for 115 countries from 1947 to 1996. We use only the 'high quality' data for the period 1960-1990. The 'high quality' label is provided by Deininger and Squire on the basis of three criteria: data are (*i*) based on a national household survey, (*ii*) which is representative of the population, and (*iii*) in which all sources of income have been counted. The total number of observations in the high quality sample is 693. The data contain missing values due to limitations to the time period of data availability, and due to missing observations within that time period. For virtually all countries, data are available only every two or five years or at irregular intervals. We construct data for 5 year intervals (although we end up using 10 year intervals for almost all estimates in this paper) from 1960 to 1995 by linear inter- and extrapolation.[12] This method yields a dataset containing 370 observations for 98 countries. For 58 countries we have three or more observations. We calculated the variance of log income from the Gini coefficients, assuming that log income is distributed normally (see appendix A). Table III summarizes the main variables in the combined dataset.[13]

3.5 Estimates of the private return from inequality data

Tables IV and V present estimates of equation (35). Since the data on income inequality are not fully comparable across countries and time, we control for differences in the way the Gini is calculated. In particular, some countries use income and others expenditure data, some refer to households, others to individuals, and some use gross and others net income. We include dummy variables in the regressions to control for these differences in definitions.[14] We simplified equation (35) by dropping the time variance of the parameter $\theta_{1t} = \theta_1$. Allowing for a time varying parameter does not make much difference, but reduces the precision of the estimates.[15]

The estimates in table IV ignore country specific fixed effects. In columns (2) through (5) we present estimates of the basic model. Column (2) estimates equation (35) by OLS, column (3) is the efficient GLS estimator in the presence of heteroskedasticity across countries and columns (4) and (5) weigh countries by log GDP per worker and by log population size. The coefficient estimates are quite robust across columns. Like in table II, heteroskedasticity does not seem to be much of an issue. The coefficients on S_t and S_t^2 are always significant at the 5% level.[16] The restriction that the long and short run

[12] For interpolation we use $\widehat{x}_t = \frac{p}{p+q}x_{t-q} + \frac{q}{p+q}x_{t+p}$, where p is the time span to the next observation and $q \leq 2$ is the time span from the previous observation. For extrapolation we use the observation that is closest by. This procedure is efficient if the Gini follows a random walk, as is almost true empirically.

[13] The data are available at http://www.princeton.edu/~tvanrens/educ.

[14] Because inequality is not only the dependent variable but we include lagged inequality as a regressor as well, we include both contemporaneous and lagged dummies. To gain efficiency, we tried restricting the coefficient on the lagged dummies to the coefficient on lagged inequality times the coefficient on the contemporaneous dummies, but the difference in the results is negligible.

[15] The estimate for the coefficient on education in column (7) in table V for instance, is -0.128 with a t-statistic of 2.68 (instead of -0.117 with a t of 3.22) if we add time variation in that coefficient. The interaction terms for 1960, 1970 and 1980 are 0.01487 (0.61), 0.00222 (0.14) and 0.01224 (1.15).

[16] The coefficients on lagged education are not individually significant, but jointly they are. This is

returns are equal can never be rejected. In column (6) we explore whether these results may be driven by poor comparability of the inequality data across countries. As pointed out by Atkinson and Brandolini [1999], additive dummy variables may be insufficient to control for changes in definitions of the Gini coefficient. We therefore dummied all observations with a definitional change separately. Although most estimates are no longer significant, all coefficients still have the expected sign and order of magnitude.

Table V investigates the relevance of fixed effects. Column (1) reproduces the baseline model from table IV. In the presence of fixed effects, this estimator overestimates the coefficient of the lagged dependent variable D_{t-1} since this variable is correlated with the country specific fixed effect. Column (2) presents a regression on deviations from group means. This estimator is inconsistent and underestimates the coefficient of the lagged dependent variable since D_{t-1} is negatively correlated with the differenced error term. Columns (3) and (4) present the consistent GMM and system GMM estimators proposed by Arellano and Bond [1991] and Blundell and Bond [1998]. Comparing column (3) with column (2), the coefficient of lagged inequality goes down, suggesting that the data do not contain enough information for reliable estimation of this model. Column (4) does better in that respect but we remain suspicious about estimates based on only 55 observations (32 for the equation in first differences). None of the dynamic models reveals significant differences between the long and the short run private return. Hence, the data do not support Acemoglu's [2000, p.38] hypothesis that an increase in the supply of human capital induces so much skill biased technological change that the long run effect on the return to education is positive. Contrary to this hypothesis we find that the long run effect of a rise in the average education level is not an increase, but a compression of wage differentials. The equality of the long run and the short run private return to education can be consistent with both the basic and the extended version of our model, the latter if the upward pressure due to the increase in the stock of knowledge exactly offsets the downward pressure due to the decline in investment in new knowledge.

Since there is no evidence for dynamic effects, estimating the static model using both the first difference estimator and the usual within estimator are consistent. These estimates are presented in columns (5) and (6). Dropping the lagged variables increases the number of observations and hence the reliability of the parameter estimates substantially. To further increase efficiency, we present GLS estimates in column (7). These estimates are consistent and efficient if the country-specific effects in inequality are uncorrelated with the education level. This assumption is clearly rejected, as can be seen from the p-value of the Hausman test. However, this rejection is due to the definition dummies, which are obviously correlated with the fixed effects. This is documented in column (8), where we again present the GLS estimates, this time excluding the dummies. Then, the Hausman test does not reject the null that the GLS estimates are consistent.

Since the estimates in column (7) are the most efficient, we take this regression as our

illustrated in column (1) where we include only the first order terms for education. Note that joint significance of θ_1 and θ_2 is sufficient evidence for $\alpha_2 > 0$, since neither S_t nor S_t^2 would have any effect on income dispersion if $\alpha_2 = 0$.

benchmark in the subsequent discussion. The calculation of the private return to education from these estimation results requires information on the variance of the education distribution V, the correlation between education and other worker characteristics ρ, and the variance of those characteristics σ^2. An estimate for V can be found in table III: $V \cong 13.7$. Since we do not have reliable estimates for ρ and σ^2, the subsequent calculations are based on $\rho = 0$.[17] Then, by equation (35) the estimates in column (7) imply

$$\begin{aligned}\alpha_2 &= \sqrt{\theta_2/V} = 0.015\\ \alpha_1 &= \frac{\theta_1}{2\alpha_1 V} = \frac{\theta_1}{2\sqrt{\theta_2 V}} = 0.28\end{aligned}$$

These parameter values imply a private return at the average education level of 5.3 years in 1990 of 20% and a compression elasticity of $\gamma = 0.37$ by equations (30) and (31). For the US where the average education level is 12 years, the implied private return is 10% and the value for the compression elasticity $\gamma = 1.50$, very close to the value of 2 implied by Katz and Murphy's [1992] estimates for the United States. Because we do not find any evidence of a dynamic pattern in the response of the private rate of return to a shock in the mean level of education, the estimates in table V are comparable to the cross section evidence in presented in table II. The estimated returns to education implied by table V are somewhat higher than those in table II, particularly for countries with a low level of education, but it is encouraging that both estimates yield comparable numbers. The value for the compression elasticity implied by Katz and Murphy's estimates lies between our estimates of 3.44 based on data on returns in table II and 1.5 based on inequality data in table V.

The effect of the variance of the education distribution on income inequality is insignificant and the coefficient estimate is unstable across the various specifications. This suggests that the direct effect of schooling on the income distribution (a more homogeneous human capital distribution leads to less income dispersion) is less important than the indirect, general equilibrium effect (a higher average education level reduces the return to human capital and therefore compresses the income distribution).[18] However, since we only have a rough proxy for V_t, the low coefficient might also be due to attenuation bias.

In table VI we further explore the robustness of the estimates for outliers, by sequencially excluding 10 different countries from the sample. This does not affect the results very much, which is remarkable given the small sample size and notoriously noisy data

[17] This provides a lower bound on the effect of education on wage dispersion

$$\theta_{1t} = 2\alpha_1\alpha_2 V + 2\alpha_2 V^{1/2}\sigma\rho = 2\alpha_2 V\left(\alpha_1 + V^{-1/2}\sigma\rho\right) \geq 2\alpha_1\alpha_2 V$$

An upper bound can be found by setting $\rho = 1$ and σ^2 equal to the total variance of log wages: $\sigma = D_t^{1/2} \simeq 0.75$ from table III. In that case $V^{-1/2}\sigma\rho \simeq 0.2$, a bit smaller than the estimate for α_1. Hence, setting $\rho = 0$ will not greatly affect the conclusions in the text.

[18] In the context of the model this finding makes sense: the sign of the effect of V_t on D_t is ambiguous, since a fall in V_t raises w_t' (see Teulings [2002]). Whether the direct or the indirect effect dominates depends on model parameters.

on inequality. All the relevant coefficients stay very close to their baseline estimates and typically remain significant. The ratio between the first and second order effects, and therefore the implied return to schooling, is even more robust than the level of the estimates.

Summarizing, our analysis of the private rate of return to education based on inequality data leads to the following conclusions: (*i*) there is strong evidence for imperfect substitution between workers of different education levels, (*ii*) the compression elasticity is in line with Katz and Murphy's [1992] estimate for the US, (*iii*) the implied private return is reasonably consistent with the evidence from microdata in different countries, in particular for countries with a high level of education, and (*iv*) the long run effect of an increase in the average education level on the private return is approximately equal to the short run effect and negative, contradicting Acemoglu's [2002] hypothesis of overshooting.

3.6 Estimates of the social return from GDP data

Estimation results for equation (36) are reported in table VII. Column (1) replicates Krueger and Lindahl [2000, table 3]. The results differ slightly because we use GDP per worker rather than GDP per capita. The short run social return to education of 8% is roughly consistent with estimates of the private return. The long run effect takes a long time to materialize, as can be seen from the coefficient of lagged GDP which is close to 1, but is 6 times larger than the short run effect and exceeds by far any estimate of the Mincerian rate of return to schooling.

Column (2) adds the crucial second order effect in education. Its coefficient has the expected negative sign and is significant at the 5% level. Again, the long run social return to education exceeds the short run return by about a factor 6, suggesting that endogenous technological progress is important to explain the long run effects of education on GDP.

When we allow for skill biased technological change in column (3) the coefficient on education and the implied social return to education seem to increase substantially, but this is because the reference category for the time dummy interactions is 1990, the last period in our sample, and therefore applying the most skill biased technology. The short run cross-effects of time dummies and education are not very precisely measured, but they are jointly significantly negative. This is clear evidence of skill biased technological progress.

We report some specification tests in columns (4) through (6). Column (4) adds the variance in the years of education. This does not affect the results. In columns (5) and (6) observations are weighed by log GDP per worker and log population size respectively to check the importance of heteroskedasticity. Again, this does not make much difference.

The regressions in table VIII control for fixed effects. Column (1) replicates column (3) of table VII. Columns (2) and (3) present the within and first difference estimators, which are inconsistent and underestimate the coefficient on the lagged dependent variable. Columns (4) and (5) present the GMM and system GMM estimators. All coefficients in column (4) are insignificant, as was to be expected given the short time dimension of our data, but have the right sign and order of magnitude. The system estimator dramatically improves the efficiency of the estimates, and the Sargan test does not reject the validity of the instruments. Hence, we take these results as the benchmark for our discussion.

The estimation results suggest the presence of country specific fixed effects. This does not come as a surprise. Gallup, Sachs and Mellinger [1999] argue for the importance of geography for growth and GDP. Access to open sea or navigable rivers is an important advantage. Countries with a temperate climate do much better than countries in the tropical zone. Acemoglu, Johnson and Robinson [2001] argue that country specific differences in growth are largely due to institutional differences such as the protection of property rights. In both cases the effect of these characteristics is largely fixed over time and therefore likely to be correlated to the level of S_t due to reverse causality, leading to overestimation of the long run effect of education. The estimation results confirm this conclusion. The long run social return drops by some 30 percentage points from column (1) to column (5) in table VIII.

At first sight, the estimate of $\beta_1 = 0.38$ for the social return seems to be substantially higher than $\alpha_1 = 0.28$ for the private return. However, the estimated short run social return is a mixture of the true short run social return and the true long run social return. Since the long run return is much higher than the short run return, this yields an overestimation of the short run return.[19] Correcting for this bias yields $\beta_1 = 0.32$, which is very close to the estimate for α_1. For countries with low education levels, the short run private and social return to education are approximately equal.

For higher education levels, the short run social return is substantially lower than the private return, since $\beta_2 = 0.04 > \alpha_2$: the negative second order effect is much stronger for the social than for the private return. This is in line with the prediction of the model that the short run private return should exceed the short run social return due to the non-tangible costs of investments in new knowledge, see section 2.2.3. These costs seem to be higher in countries with high education levels. It is tempting to relate this result to global externalities in knowledge production: the low S_t countries do not invest much in knowledge themselves, but free ride on the investments of the high S_t countries. However, high S_t countries have little to complain, because the skill bias in the newly invented technologies shifts technology in favor of their highly educated workforce.

The long run social return is two times higher than the private return at the average education level in 1990 of 5.3 years. This is clear evidence in favor of the extended model with endogenous technological progress. However, the estimate of ψ suggests that it takes several decades before a substantial part of the surplus of the long above the short run return is actually realized. The estimation results confirm another prediction of the extended model, namely that the social return rises more steeply than the private

[19] Let β_{LR} be the true long run effect, β_{SR} the true short run effect, and β the estimated short run effect. If a a shock to S hits just before the beginning of an observation period, the short run effect is estimated correctly. However, if a shock hits just after the start of an observation period, the estimated short run effect is $\frac{1}{2}\left((1+\psi)\,\beta_{\mathrm{SR}} + (1-\psi)\beta_{\mathrm{LR}}\right)$. Averaging over both extremes gives $\beta_{\mathrm{SR}} = (4\beta - (1-\psi)\,\beta_{\mathrm{LR}})\,/\,(3+\psi)$.

return.

The results in column (5) show very strong (and precisely estimated) second order effects of education on GDP, both in the short and in the long run. These estimates indicate that both the short and the long run social return are strongly declining in the average eduction level of the workforce: the implied return is $0.38 - 0.04S$ in the short run, and $1.07 - 0.12S$ in the long run.[20] Hence, GDP growth of low S_t countries is mainly due to the high social rate of return on their investments in a better educated workforce, while GDP growth of high S_t countries is mainly due to the skill bias in technological progress, which shifts technology to their advantage.

The interaction terms of education with time dummies indicate clear evidence for skill biased technological progress, raising the short run social return to education at constant S_t by about 4% during the seventies and by 3% during the eighties. The coefficient of the second order term in education equals $\frac{1}{2}\beta_2$, so to offset the 3% increase in the return during the 80s, the average education level would have to increase by $3\%/\left(2 \cdot 2\%\right) = 0.75$ years. The effect of skill biased technological progress on the return to schooling was therefore about the same size as the effect of the increase in the average education level over the same period (the average education level in our sample increased by 0.8 years in the 70s and by 0.7 years in the 80s, see table III). The race between education and technology has no clear winner: the upward effect of technology is offset by the increase in the average education level across the world. A combination of Tinbergen's race between education and technology, endogenous technological progress and country-specific fixed effects describes the evolution of GDP between 1960 and 1990 fairly well.

Krueger and Lindahl [2000] have shown that estimates of the return to human capital from this type of model are strongly affected by attenuation bias because of measurement error when using short time intervals. However, the longer the time interval, the greater the risk of reverse causality. As a compromise we use 10 year time periods. In table IX we report some sensitivity analysis for this choice. Columns (2), (4), (6) and (8) present Krueger and Lindahl's specification and columns (1), (3), (5) and (7) our preferred specification, using 5, 10 and 20 year time periods. Reading the table horizontally, we see that the coefficient estimates for the short run return to education increase as we use longer time intervals. Part of this might be due to the same problem of aggregation in time that we discussed before, see the footnote on page 26, which is problematic for the conclusions of Krueger and Lindahl. Because the long run return is substantially higher than the short run return, we can increase the estimate of the short run return to almost any level by using longer and longer time intervals. Columns (7) and (8) repeat the estimation for 20 year time intervals with the Kyriacou [1991] data for education. The results are largely similar to those using the Barro and Lee education data.

Table X presents further robustness checks. Our results might be driven by a few countries with exceptionally high growth rates and exceptionally high investment in human capital, both persisting over the whole 30 year period covered. This would open a channel for reverse causality by the following story: some countries grow fast over

[20] In fact the second order effects are so large that the implied return to education becomes negative at education levels of over 10 years. Clearly, this is an artefact of the linearization of (30) and (37).

prolonged period, and use their additional revenues to invest in education. In that case, the increase in the average level of education in this observation period is just a predictor of the raise in education during the previous observation period. Hence, we exclude the 10 highest and lowest observations on $\Delta y_t, \Delta S_t, y_t$ and S_t in a number of regressions. The estimates are quite robust to this procedure.

Concluding, there is (i) strong evidence for imperfect substitution between workers of different education levels also for the social rate of return to education measured from data on GDP, (ii) the short run social return to education approximately equals the private return, and is lower than the private return for countries with high education levels as predicted by our model, (iii) the long run social return is two times higher than the private return, even after controlling for fixed effects, providing strong evidence for endogenous technological progress, and (iv) there is evidence for skill biased technological change and taking the time variation in the social return to education into account is important to describe the evolution of GDP from 1960 to 1990.

4 Conclusions

We have captured the evolution of the social and the private rate of return to education by a simple model of imperfect substitution between workers with various levels of education and endogenous skill-biased technological progress. Human capital enters as a factor of production in this simple constant returns to scale Cobb-Douglas economy. In the short run, the Walrasian equality between the private and the social return to education applies. In the long run, an increase in the average education level of the workforce induces investment in new knowledge, which leads to skill biased technological progress. This pushes the long run social rate of return to education above the long run private rate. We derived easy to interpret relations between educational attainment, GDP and the social rate of return, and between educational attainment, income inequality and the private rate of return, which we estimated from cross-country panel data. Our empirical results provide strong support for the negative relation between the supply of human capital and its private and social return. The estimates imply that a one year increase in the stock of human capital reduces its return by 1.5 (for the private return) to 4 (for the social return) percentage points. The estimate for the private return is well in line with conventional estimates of the elasticity of substitution between low and high skilled workers, cf. Katz and Murphy [1992].

The short run social return to education is approximately equal to the private return. For countries with high education levels, the costs in terms of output of non-tangible investments in knowledge further decrease the short run social return, and for those countries the short run social return is lower than the private return, as predicted by our theoretical model. This result suggests that countries with a highly educated workforce invest more in new technologies. This finding might be due to global externalities in knowledge production. Countries with a low average education level free ride on the investments of the highly educated countries. However, the latter have little to complain, because the skill bias in new knowledge shifts technology in favor of their highly educated workforce. Since we use cross-country data, we cannot estimate these global externalities; in our regressions they are absorbed by the time dummies.

The long run social return is about two times higher than the private return. Allowing for country-specific fixed effects in GDP matters substantially here, as it brings down the estimated long run social return evaluated at the average education level from 70% to around 40%.[21] The intuition is that there is reverse causality from GDP to the average education level: countries with favorable fixed effects in GDP have higher levels of education. Not accounting for these fixed effects will therefore overestimate the effect of education on GDP. Both the long and short run social return are low, or even negative for countries with a highly educated workforce. Hence, lowly educated countries grow mainly due to the high social rate of return to their investment in a better educated workforce, while highly educated countries grow mainly due to the skill bias in new knowledge, which shifts technology to their advantage.

Our estimates for the GDP equation represent a substantial improvement over the existing growth literature, which tends to find no effect of increases in education on GDP growth. Partly, this is due to measurement error, as argued by Krueger and Lindahl [2000]. However, their estimates yield long run returns that are six times higher than the short run return. We add a second order term to account for a general equilibrium effect through imperfect substitution between low and high educated workers and allow for skill biased technological change. Both extensions help to decrease the gap between the estimates of the social and private return. We also show that it is important to allow for the long run and short run return to differ. Krueger and Lindahl's finding that the estimated social return increases with the time intervals used, which they attribute to measurement error, may be driven in part by the fact that the long run social return is substantially higher than the short run return.

Allowing the effect of education on GDP to vary over time, we find clear evidence for skill biased technological progress.[22] Our estimates do not show an accelleration of skill biased technological progress during the 1980s as the literature has suggested. The observed skill biased technological progress is endogenous, which may explain why the long run social return to education is so high. However, this endogenous skill biased technological progress cannot have been responsible for the increased inequality in the 1980s in the US, as suggested by Acemoglu [2002]. Acemoglu argues that an increase in the average level of education may induce so much skill biased technological progress that the initial negative effect on the private return to education gets reversed. Theoretically, our model allows for this kind of overshooting of the return to human capital in response to a shock to its supply. However, empirically we do not find support for this implication:

[21] A long run return to education of 40% may still seem high, but the reader should keep in mind that our regressions do not include capital as an explanatory variable, since it is correlated with GDP by construction. Hence, we are unable to separate the effect of human and physical capital on GDP. By implication, our effect of human capital includes the induced effect of human capital accumulation on the accumulation of physical capital.

[22] And possibly for externalities in the production of new technologies. The effects of externalities and technological progress cannot be distinguished from eachother from the estimated long run private and social returns to education because $\bar{y}' = \bar{w}' + \frac{\rho}{\theta}\bar{P}_x\bar{X}'$, see equation (24). In principle we could use the estimated speed of convergence to disentangle the two effects, but these estimates are not very reliable.

a larger supply of human capital reduces the private return to education unambiguously.

How do our estimation results fit into the global picture of post war productivity growth? From 1970 to 1990, the average education level of the countries in our sample increased by about 0.066 years per year. The long run social return, evaluated in 1990 the average education level of 5.3 years is about 40%, which implies a 2.7% productivity growth per year. Actual productivity growth is only 2% per year over the sample period (see table III), suggesting that our model yields *more* productivity growth than is actually observed in the data. There are two explanations for this discrepancy. Since the long run social return takes several decades to realize, the full effect of the post war increase in education on productivity may not yet have been realized in the sample period. Secondly, skill biased technological progress shifts the terms of trade at a global level. This could make countries that do not increase the education level of their work force worse off. This explanation is confirmed by our estimates: the coefficient of the time dummies show that a country that would not have increased its education level would have experienced a *decrease* in GDP from 1970 to 1990. This finding complicates traditional growth accounting, since the effect of education on GDP can account for more than 100% productivity growth. In any case it is clear that education is an enormously important factor in explaining post war differences in GDP growth both between countries and over time.

A Gini coefficient and the variance of log income

Let $W \in \left[\underline{W}, \overline{W}\right]$ denote income with density $f(W)$, distribution function $F(W)$ and mean M. $F(W)$ measures the share of the population with income lower than W. Let $Z(W)$ denote the cumulative share of total income earned by people with income lower than W. By definition:

$$Z(W) = \frac{1}{M} \int_{\underline{W}}^{W} x f(x) dx \tag{39}$$

The graph of the Lorenz curve has $F(W)$ on the horizontal and $Z(W)$ on the vertical axis. The Gini coefficient $G \in [0, 1]$ is given by twice the area between the Lorentz curve and the 45-degree line.

$$G = 1 - 2 \int_0^1 Z dF = 2 \int_0^1 F dZ - 1$$

By change of variables, using $dZ = \frac{1}{M} W f(W) dW$, this expression can be written as:

$$G = \frac{2}{M} \int_{\underline{W}}^{\overline{W}} W f(W) F(W) dW - 1$$

Assume income to be log normally distributed so that $F(W) = \Phi\left(\frac{w-\mu}{\sigma}\right)$ and $M = e^{\mu+\frac{1}{2}\sigma^2}$, where $w \equiv \ln W$ and μ and σ^2 are the mean and variance of w. By change of variables $v = \frac{w-\mu}{\sigma} \Rightarrow dW = \sigma e^{\sigma v+\mu} dv$, the Gini coefficient can written as:

$$G = \frac{2}{M} \int_0^\infty W \frac{\phi\left(\frac{w-\mu}{\sigma}\right)}{\sigma W} \Phi\left(\frac{w-\mu}{\sigma}\right) dW - 1 = 2e^{-\frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} e^{\sigma v} \phi(v) \Phi(v) \, dv - 1$$

which maps the Gini coefficient to the variance of the log income distribution σ^2. Numerically evaluating this expression for different values of σ shows that the relationship is virtually linear in the relevant range. Variances of log income of 0, 0.1, 0.2, 0.3 and 0.4 correspond to Gini coefficients of 52.05, 56.33, 60.39, 64.20 and 67.78 respectively.

References

[1] Acemoglu, Daron (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature,* vol.40 no.1, pp.7-72.

[2] Acemoglu, Daron and Joshua Angrist (1999). How large are Social Returns to Education? Evidence from Compulsory Schooling Laws. NBER Working Paper No.7444.

[3] Acemoglu, Daron, Simon Johnson and James A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review,* vol.91 no.5, pp.1369-1401.

[4] Arellano, Manuel and Steve Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies,* vol.58 no.2, pp. 277-297.

[5] Atkinson A.B. and A. Brandolini (1999). *Promise and Pitfalls in the Use of "Secondary" Data-Sets: Income Inequality in OECD Countries.* mimeo, available at: http://www.nuff.ox.ac.uk/economics/people/atkinson.htm or: http://www.bancaditalia.it/pubblicazioni/temidi

[6] Barro, Robert J. and Xavier Sala-i-Martin (1999). *Economic Growth.* Cambridge MA: MIT Press (first MIT Press edition, originally published by McGraw-Hill, 1995).

[7] Barro, Robert J. and Jong Wha Lee (1993). International Comparisons of Educational Attainment. *Journal of Monetary Economics,* vol.32 no.3, pp.363-394 (dataset available at: http://www.worldbank.org/research/growth/ddbarlee.htm).

[8] Barro, Robert J. and Jong Wha Lee (1996). International Measures of Schooling Years and Schooling Quality. *American Economic Review,* vol.86 issue 2 Papers and Proceedings May 1996,

pp.218-223 (dataset available at: http://www.nber.org/data or http://www.worldbank.org/research/growth/ddbarle2.htm).

[9] Benhabib, Jess and M. Spiegel (1994). The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data. *Journal of Monetary Economics,* vol.34 no.2, pp.143-174.

[10] Bils, Mark and Peter J. Klenow (1998) Does Schooling Cause Growth or the Other Way Around? *NBER Working Paper* No.6393.

[11] Blundell, Richard and Stephen Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics,* vol. 87, pp.115-143.

[12] Card, David (1999). The Causal Effect of Education on Earnings. Chapter 30 in the *Handbook of Labor Economics,* Volume 3, O. Ashenfelter and D. Card (eds.). Amsterdam: Elsevier.

[13] Caselli, Francesco and Wilbur John Coleman (2002). *American Economic Review,* vol.92, iss.2, pp.148-52.

[14] Checchi, Daniele (1999). *Does Educational Achievement Help to Explain Income Inequality?* mimeo, Università degli Studi di Milano Bicocca. Available at http://www.eco-dip.unimi.it/pag_pers/checchi/checchi1.htm.

[15] Checchi, Daniele and Luca Flabbi (1999). *Income and Educational Distribution Dataset: Various Countries in Panel Format* (description of the dataset, data available at: http://www.eco-dip.unimi.it/pag_pers/checchi/checchi1.htm)

[16] Deininger, Klaus and Lyn Squire (1996). A New Data Set Measuring Income Inequality. *World Bank Economic Review,* vol.10 no.3, pp.565-591 (dataset available at: http://www.worldbank.org/research/growth/dddeisqu.htm).

[17] Doornik, Jurgen A., Manuel Arellano and Stephen Bond (2002). Panel Data estimation using DPD for Ox. Available at http://www.nuff.ox.ac.uk/Users/Doornik/

[18] Gallup, J.L., J.D. Sachs and A.D. Mellinger (1999). Geography and economic development. *NBER Working Paper* No.6849.

[19] Heckman, James J. and Peter J. Klenow (1997). *Human Capital Policy.* mimeo, University of Chicago. Available at: http://www.klenow.com/.

[20] Katz, Lawrence F. and Kevin M. Murphy (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics,* vol.107 issue 1, pp.35-78.

[21] Krueger, Alan B. and Mikael Lindahl (2000). Education for Growth: Why and For Whom? *NBER Working Paper* No.7591. Forthcoming in *Journal of Economic Literature,* vol.39 nr.4 (Dec 2001).

[22] Kyriacou, George (1991). Level and Growth Effects of Human Capital: A Cross-Country Study of the Convergence Hypothesis. *New York University C.V. Starr Center for Applied Economics Working Paper* RR 9126 (www.econ.nyu.edu/working) (dataset available on request).

[23] Lucas, Robert. E (1988).On the Mechanisms of Development Planning, *Journal of Monetary Economics,* vol.22 no.1, pp.3-42.

[24] O'Neill, Donald (1995). Education and Income Growth: Implications for Cross-Country Inequality. *Journal of Political Economy,* vol.103 no.6, pp.1289-1301.

[25] Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development,* vol.22 no.9, pp.1325-1343.

[26] Summers, Robert and Alan Heston (1991). The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988. *Quarterly Journal of Economics,* vol.106 no.2, pp. 327-368 (the new dataset, mark 5.6a, is available at: http://pwt.econ.upenn.edu or http://www.nber.org/data).

[27] Teulings, Coen N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy,* vol.103 no.2, pp.280-315.

[28] Teulings, Coen N. (2002). *Comparative Advantage, Relative Wages, and the Accumulation of Human Capital.* Discussion paper TI 2002-081/3, Tinbergen Institute, Amsterdam/Rotterdam

[29] Tinbergen, Jan (1975). *Income Distribution: Analysis and Policies.* Amsterdam: North Holland.

[30] Uzawa, Hirofumi (1965). Optimal Technical Change in an Aggregative Model of Economic Growth, *International Economic Review*, 6, pp.18-31.

Table I.
Return to Schooling Across Countries

PWT 5.0 country code	Country		Average years of schooling population over 25: year	educ. level	Return to Education: year	ret. to educ
123	Poland	POL	85	8.7	86	.024
126	Sweden	SWE	80	9.45	81	.026
114	Greece	GRC	85	6.89	85	.027
118	Italy	ITA	85	5.75	87	.028
107	Austria	AUT	85	7.17	87	.039
115	Hungary	HUN	85	7.93	87	.039
50	Canada	CAN	80	10.23	81	.042
83	China	CHN	85	4.04	85	.045
110	Denmark	DNK	90	11.21	90	.047
89	Israel	ISR	80	9.11	79	.057
85	India	IND	80	2.72	81	.062
131	Australia	AUS	80	10.02	82	.064
121	Netherlands	NLD	85	8.29	83	.066
41	Tanzania	TZA	80	.	80	.067
127	Switzerland	CHE	85	8.99	87	.072
68	Bolivia	BOL	90	4.11	89	.073
113	Germany West	DEU	90	8.83	88	.077
53	Dom. Rep.	DOM	90	3.76	89	.078
117	Ireland	IRL	85	7.87	87	.079
78	Venezuela	VEN	90	4.89	89	.084
75	Peru	PER	90	5.5	90	.085
21	Kenya	KEN	80	2.46	80	.085
77	Uruguay	URY	90	6.69	89	.09
104	Thailand	THA	70	3.54	71	.091
66	USA	USA	90	12	89	.093
94	Malaysia	MYS	80	4.49	79	.094
124	Portugal	PRT	85	3.45	85	.094
29	Morocco	MAR	70	.	70	.095
54	El Salvador	SLV	90	3.4	90	.096
129	UK	GBR	70	7.66	72	.097
97	Pakistan	PAK	80	1.74	79	.097
61	Nicaragua	NIC	80	2.83	78	.097
109	Cyprus	CYP	85	7.56	84	.098
72	Ecuador	ECU	85	5.36	87	.098
74	Paraguay	PRY	90	4.72	89	.103
51	Costa Rica	CRI	90	5.4	89	.105
92	Korea	KOR	85	8.03	86	.106
67	Argentina	ARG	90	7.77	89	.107
100	Singapore	SGP	75	4.38	74	.113
98	Philippines	PHL	90	6.73	88	.119
70	Chile	CHL	90	6.16	89	.121
4	Botswana	BWA	80	2.29	79	.126
62	Panama	PAN	90	7.55	89	.126
125	Spain	ESP	90	6.25	90	.13
60	Mexico	MEX	85	4.34	84	.141
56	Guatemala	GTM	90	2.56	89	.142
71	Colombia	COL	90	4.25	89	.145
69	Brazil	BRA	90	3.56	89	.154
86	Indonesia	IDN	80	3.09	81	.17
58	Honduras	HND	90	3.68	89	.172
20	Cote d'Ivoire	CIV	85	.	85	.207
59	Jamaica	JAM	90	4.51	89	.28

Education data from Barro and Lee. Return to education data from Bils and Klenow (1998).
Original sources return to education: Rosholm and Smith 1996 (Denmark), Calan and Reilly 1993 (Ireland), Armitage and Sabot 1987 (Kenya and Tanzania), Alba-Ramirez and San Segundo 1995 (Spain), Arai 1994 (Sweden), Chiswick 1977 (Thailand), Krueger and Pischke 1992 (USA and Germany) and Psacharopoulos 1994 (all other countries); see Bils and Klenow for full references.

Table II.
Direct Estimates of Diminishing Returns to Schooling

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS excl. Jamaica	WLS (GDP/w)	WLS (GDP/w) excl. Jamaica	WLS (population)	WLS (population) excl. Jamaica
S_t	-0.00708	-0.00638	-0.00721	-0.00649	-0.00673	-0.00614
	(3.23)	(3.68)	(3.41)	(3.86)	(3.18)	(3.49)
(year=70)	-0.02297	-0.01538	-0.02100	-0.01382	-0.02247	-0.01620
	(0.81)	(0.69)	(0.75)	(0.62)	(0.85)	(0.74)
(year=80)	-0.03538	-0.02759	-0.03542	-0.02819	-0.03556	-0.02902
	(2.49)	(2.44)	(2.52)	(2.51)	(2.55)	(2.49)
(year=85)	-0.04061	-0.03381	-0.04012	-0.03365	-0.04270	-0.03700
	(3.06)	(3.21)	(3.13)	(3.28)	(3.30)	(3.42)
Constant	0.15663	0.14513	0.15725	0.14591	0.15451	0.14490
	(10.33)	(11.95)	(10.50)	(12.07)	(10.34)	(11.54)
Observations	49	48	49	48	49	48
R-squared	0.36	0.40	0.37	0.41	0.36	0.39

t statistics in parentheses. Dependent variable is the return to education as in table II. WLS regressions are weighted by log GDP per worker or log population size. The dummy for 1975 was dropped because there is only one observation in that year.

Table III.
Summary Statistics

		1970	1980	1990	full sample	Description and source
y_t	mean	8.5611	8.8598	8.8617	8.7585	Log real GDP per worker, 1985 intl prices, Chain index (PWT 5.6a)
	sd	1.0249	1.0842	1.0505	1.0612	
	obs	133	142	115	390	
Δy_t	mean	0.0325	0.0226	0.0020	0.0198	10 year changes in real GDP per worker (annualized).
	sd	0.0226	0.0264	0.0249	0.0276	
	obs	125	133	111	369	
D_t	mean	0.6687	0.5148	0.5631	0.5727	Variance of log income. Calculated from Gini coefficient (Deininger and Squire).
	sd	0.3912	0.2701	0.3306	0.3318	
	obs	39	52	77	168	
ΔD_t	mean	-0.0042	-0.0026	0.0032	0.0000	10 year changes in variance log income (annualized).
	sd	0.0178	0.0168	0.0164	0.0169	
	obs	15	29	45	89	
S_t	mean	3.8272	4.5635	5.3194	4.5766	Average years of education attained by population over 25 years old (Barro and Lee).
	sd	2.7660	2.9402	2.9445	2.9408	
	obs	109	113	112	334	
ΔS_t	mean	0.0414	0.0826	0.0741	0.0662	10 year changes in average years of education (annualized).
	sd	0.0640	0.0669	0.0614	0.0663	
	obs	107	109	112	328	
V_t	mean	11.2290	13.4041	16.1395	13.6956	Variance of the education distribution (proxy constructed from Barro and Lee data).
	sd	5.0021	5.5804	6.2650	5.9895	
	obs	90	102	103	295	
ΔV_t	mean	0.1688	0.2842	0.2814	0.2489	10 year changes in variance education (annualized).
	sd	0.2771	0.3217	0.2787	0.2966	
	obs	81	90	102	273	

Table IV.
Income Inequality: OLS Estimates

	(1) OLS	(2) OLS	(3) RE	(4) WLS (GDP/w)	(5) WLS (population)	(6) OLS (definition dummies)
S_t	-0.10997	-0.26883	-0.27796	-0.28081	-0.25649	-0.13200
	(3.40)	(2.81)	(3.02)	(2.99)	(2.72)	(1.15)
S_t^2		0.01177	0.01248	0.01238	0.01116	0.00546
		(2.02)	(2.02)	(2.17)	(1.97)	(0.81)
V_t	0.00057	0.00813	0.01019	0.00854	0.00714	0.00606
	(0.09)	(1.19)	(1.38)	(1.26)	(1.04)	(0.72)
LR: S	-0.08268	-0.33596	-0.28604	-0.33533	-0.37230	-1.33958
	(2.62)	(1.57)	(1.53)	(1.16)	(1.64)	(0.38)
LR: S^2		0.01597	0.01300	0.01593	0.01836	0.08785
		(1.13)	(1.00)	(1.61)	(1.23)	(0.36)
S_{t-1}	0.08948	0.18955	0.19605	0.20104	0.17587	0.06795
	(3.02)	(2.06)	(2.28)	(2.22)	(1.94)	(0.66)
S_{t-1}^2		-0.00800	-0.00876	-0.00859	-0.00718	-0.00126
		(1.40)	(1.53)	(1.53)	(1.28)	(0.20)
V_{t-1}	0.00768	0.00280	0.00075	0.00277	0.00360	0.00576
	(1.09)	(0.38)	(0.10)	(0.38)	(0.48)	(0.66)
D_{t-1}	0.75213	0.76400	0.71367	0.76209	0.78345	0.95219
	(7.51)	(7.78)	(7.23)	(7.71)	(8.36)	(7.28)
(year=70)	-0.08903	-0.07420	-0.06949	-0.07208	-0.06840	-0.01736
	(1.62)	(1.41)	(1.15)	(1.38)	(1.37)	(0.31)
(year=80)	-0.03914	-0.03732	-0.03996	-0.03520	-0.03275	-0.01494
	(0.80)	(0.74)	(0.88)	(0.71)	(0.67)	(0.27)
Constant	0.19901	0.32954	0.35559	0.33459	0.31479	0.11402
	(1.57)	(2.06)	(2.32)	(2.10)	(2.10)	(0.65)
Observations	77	77	77	77	77	77
R-squared	0.82	0.83		0.83	0.85	0.91
SR return[a]		0.17948	0.17613	0.18166	0.17672	0.13546
		(4.40)	(3.87)	(3.02)	(4.34)	(1.96)
LR return[a]		0.20375	0.17888	0.19960	0.21990	0.50470
		(3.01)	(3.19)	(4.66)	(2.96)	(0.65)
LR=SR test		0.07	0.00	0.04	0.19	0.21
p-value		0.7939	0.9700	0.8418	0.6644	0.6493

Robust t-statistics in parenthesis. The first five columns include six dummies for the definition of the Gini coefficient: income-expenditure, household-individual and gross-net, both contemporaneous and lagged. The last column includes the three contemporaneous dummies, plus 21 dummies for each definition change in a country in the sample period.

a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Table V.
Income Inequality: Panel Estimates

	(1) OLS levels	(2) FE (incons.)	(3) Arellano-Bond	(4) Blundell-Bond	(5) OLS first difs	(6) FE	(7) RE	(8) RE
S_t	-0.26883	-0.21491	-0.14186	-0.49320	-0.08561	-0.08262	-0.11717	-0.09049
	(2.81)	(1.63)	(1.29)	(3.06)	(1.29)	(1.50)	(3.22)	(2.23)
S_t^2	0.01177	0.01042	0.00882	0.02700	0.00292	0.00226	0.00314	0.00346
	(2.02)	(1.20)	(1.43)	(2.85)	(0.72)	(0.58)	(1.09)	(1.07)
V_t	0.00813	-0.00117	-0.00311	0.01997	0.00275	-0.00562	-0.00428	-0.00147
	(1.19)	(0.09)	(0.31)	(2.11)	(0.39)	(0.90)	(0.93)	(0.28)
LR: S	-0.33596	-0.04072	-0.02002	-0.67446				
	(1.57)	(0.33)	(0.73)					
LR: S^2	0.01597	-0.00182	-0.00179	0.03840				
	(1.13)	(0.49)	(0.51)					
S_{t-1}	0.18955	0.15296	0.11044	0.17340				
	(2.06)	(1.54)	(1.16)	(1.42)				
$S_{t-1}{}^2$	-0.00800	-0.01319	-0.01163	-0.00879				
	(1.40)	(1.84)	(1.70)	(1.20)				
V_{t-1}	0.00280	-0.01270	-0.01275	0.00753				
	(0.38)	(1.13)	(1.22)	(0.95)				
D_{t-1}	0.76400	-0.52111	-0.56923	0.52584				
	(7.78)	(2.21)	(1.92)	(2.42)				
(year=60)						-0.16610	-0.21177	-0.06900
						(2.11)	(3.67)	(1.16)
(year=70)	-0.07420	-0.27224			-0.09725	-0.16442	-0.19096	-0.08466
	(1.41)	(1.74)			(1.83)	(2.63)	(4.26)	(1.89)
(year=80)	-0.03732	-0.20657		0.00608	-0.05407	-0.11530	-0.13483	-0.06397
	(0.74)	(2.22)		(0.09)	(1.11)	(3.22)	(4.40)	(2.05)
(year=90)			0.18057	0.05836				
			(2.29)	(0.80)				
Constant	0.32954	1.40073	0.01836	0.68270	-0.04350	0.85883	1.00626	1.03122
	(2.06)	(2.19)	(0.35)	(2.22)	(0.61)	(4.30)	(10.51)	(9.68)
Observations	77	77	32	55	77	148	148	149
Nr countries	45	45	23	23	65	65	65	66
SR return[a]	0.17948	0.13826	0.13837	0.17018	0.13652	0.16674	0.20239	0.12370
	(4.40)	(2.14)	(2.45)		(1.50)	(1.23)	(2.33)	(2.30)
LR return[a]	0.20375			0.18435				
	(3.01)							
LR=SR test[b]	0.07	3.44	2.60					
p-value	0.7939	0.9215	0.8777					
Sargan chi2			10.49	19.33				
p-value			0.487	0.781				
Hausman							119.50	1.20
p-value							0.0000	0.9768

Robust t-statistics in parenthesis. All columns except the last include dummies for the definition of the Gini coefficient: income-expenditure, household-individual and gross-net. The first four columns include both contemporaneous and lagged dummies, the last four only contemporaneous. The Arellano-Bond GMM and the Blundell-Bond system GMM estimators assume that the regressors are predetermined (not necessarily exogenous). The coefficient estimates in the table are 1-step estimates since the efficient 2-step estimates are prone to overfitting.

a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

b) The tests in columns (2) and (3) are based on first order effects only because the long run effect of education squared is negative and we therefore cannot calculate the implied long run return to education.

Table VI.
Income Inequality: Subsample Robustness

	(1) Full sample	(2) Without 10 countries with highest growth in education	(3) Without 10 countries with highest growth in inequality	(4) Without 10 countries with highest education level	(5) Without 10 countries with highest inequality	(6) Without 10 countries with lowest education level	(7) Without 10 countries with lowest inequality
S_t	-0.11717	-0.14893	-0.11663	-0.10016	-0.06248	-0.12112	-0.09351
	(3.22)	(3.84)	(3.22)	(2.04)	(1.86)	(3.08)	(2.16)
S_t^2	0.00314	0.00500	0.00434	0.00219	0.00077	0.00333	0.00203
	(1.09)	(1.65)	(1.46)	(0.49)	(0.30)	(1.12)	(0.61)
V_t	-0.00428	-0.00265	-0.00193	-0.01021	0.00163	-0.00326	-0.00993
	(0.93)	(0.50)	(0.42)	(1.86)	(0.40)	(0.73)	(1.68)
(year=60)	-0.21177	-0.17158	-0.10259	-0.25692	-0.09922	-0.21180	-0.22021
	(3.67)	(2.79)	(1.61)	(3.49)	(1.78)	(3.62)	(3.14)
(year=70)	-0.19096	-0.18536	-0.11492	-0.23495	-0.12387	-0.19508	-0.19282
	(4.26)	(3.90)	(2.37)	(4.00)	(2.93)	(4.19)	(3.35)
(year=80)	-0.13483	-0.12076	-0.08045	-0.17270	-0.08997	-0.13432	-0.15631
	(4.40)	(3.71)	(2.61)	(4.33)	(3.08)	(4.29)	(3.95)
Constant	1.00626	1.06769	0.92354	1.05523	0.64052	1.01348	1.04999
	(10.51)	(10.52)	(9.16)	(9.20)	(6.64)	(8.23)	(9.98)
Observations	148	125	120	122	132	135	120
Nr countries	65	55	55	55	55	55	55
ret to educ[a]	0.20239	0.18327	0.14496	0.22201	0.26505	0.20086	0.21594
	(2.33)	(3.58)	(2.78)	(0.94)	(0.62)	(2.53)	(1.31)
Countries excluded from the sample		Algeria Egypt Mexico Trinidad Peru Hong Kong Jordan Korea Taiwan Greece	Guatemala Brazil Chile Venezuela Hong Kong Thailand Denmark Greece Australia New Zealand	Canada USA Belgium Denmark Finland Germany Poland Sweden Australia New Zealand	C. Afr. Rep. Kenya Senegal Sierra Leone Zimbabwe Guatemala Honduras Mexico Brazil Turkey	C. Afr. Rep. Gambia Senegal Sierra Leone Uganda Zimbabwe Bangladesh Iran Pakistan Turkey	Canada India Taiwan Belgium Finland Hungary Netherlands Poland Spain UK

All columns are RE estimates (as in column 7 in table V). z-statistics in parenthesis. All regressions include three dummies for the definition of the Gini coefficient: income-expenditure, household-individual and gross-net.

a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Table VII.
GDP Equation: OLS Estimates

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) WLS (GDP/w)	(6) WLS (population)
S_t	0.08546	0.17025	0.24335	0.24508	0.24717	0.24814
	(4.22)	(3.26)	(3.86)	(2.90)	(3.82)	(4.07)
S_t^2		-0.00780	-0.00848	-0.00881	-0.00840	-0.00898
		(2.08)	(2.33)	(1.75)	(2.36)	(2.49)
S_{1970}			-0.09705	-0.07495	-0.09901	-0.09956
			(1.97)	(1.41)	(2.01)	(2.00)
S_{1980}			-0.06732	-0.07423	-0.07728	-0.06933
			(1.28)	(1.32)	(1.47)	(1.29)
V_t				-0.00461		
				(0.62)		
LR: S	0.48212	1.08877	1.45022	1.24804	1.45147	1.49990
	(4.65)	(4.16)	(4.61)	(1.72)	(3.35)	(3.36)
LR: S^2		-0.05718	-0.06873	-0.04666	-0.06786	-0.07277
		(2.71)	(3.29)	(2.92)	(4.66)	(4.55)
S_{t-1}	-0.05574	-0.08459	-0.12164	-0.15486	-0.12407	-0.12629
	(2.89)	(1.64)	(1.92)	(1.97)	(1.91)	(2.06)
S_{t-1}^2		0.00330	0.00272	0.00543	0.00265	0.00307
		(0.85)	(0.72)	(1.12)	(0.71)	(0.81)
S_{1960}			0.06210	0.04247	0.06046	0.06730
			(1.26)	(0.83)	(1.24)	(1.35)
S_{1970}			0.03732	0.03515	0.04336	0.04169
			(0.73)	(0.65)	(0.85)	(0.80)
V_{t-1}				0.00834		
				(1.25)		
y_{t-1}	0.93837	0.92133	0.91608	0.92771	0.91519	0.91877
	(44.15)	(41.71)	(41.79)	(32.88)	(41.46)	(43.20)
(year=70)	0.34489	0.35062	0.55900	0.55157	0.57686	0.54274
	(11.25)	(10.84)	(9.12)	(7.91)	(9.19)	(8.68)
(year=80)	0.21204	0.21792	0.40168	0.46588	0.42694	0.38323
	(6.34)	(6.62)	(5.65)	(5.86)	(5.84)	(5.30)
Constant	0.38155	0.40327	0.27154	0.20405	0.27347	0.26010
	(2.28)	(2.43)	(1.66)	(1.00)	(1.64)	(1.61)
Observations	292	292	292	250	292	292
R-squared	0.95	0.95	0.95	0.95	0.95	0.95
SR return[a]	0.08546	0.08755	0.15341	0.15173	0.15808	0.15297
	(4.22)	(5.49)	(3.41)	(2.93)	(5.34)	(3.43)
LR return[a]	0.48212	0.48263	0.72167	0.75345	0.73215	0.72854
	(4.65)	(4.02)	(5.36)	(3.75)	(3.44)	(5.18)

Robust t statistics in parentheses.
a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Table VIII.
GDP: Dynamic Panel Estimates

	(1) OLS	(2) FE (incons)	(3) OLS first diff (incons)	(4) Arellano-Bond	(5) Blundell-Bond
S_t	0.24335	0.16222	0.21467	0.20156	0.38124
	(3.86)	(1.88)	(2.33)	(0.65)	(4.36)
S_t^2	-0.00848	-0.00337	-0.00744	-0.02647	-0.02084
	(2.33)	(0.59)	(1.26)	(1.02)	(4.23)
S_{1970}	-0.09705	-0.07634	-0.06567	0.15145	-0.07258
	(1.97)	(1.06)	(0.92)	(0.96)	(1.20)
S_{1980}	-0.06732	-0.05327	-0.05795	0.17864	-0.03121
	(1.28)	(0.84)	(0.86)	(0.88)	(0.49)
LR: S	1.45022	0.17670	0.21839	3.74776	1.07537
	(4.61)	(0.23)	(1.51)		
LR: S^2	-0.06873	-0.00327	-0.00839	-0.27592	-0.06219
	(3.29)	(0.93)	(0.86)		
S_{t-1}	-0.12164	-0.06388	-0.02163	0.24646	-0.06142
	(1.92)	(0.66)	(0.20)	(1.04)	(0.61)
S_{t-1}^2	0.00272	0.00155	0.00002	-0.00652	0.00235
	(0.72)	(0.28)	(0.00)	(0.48)	(0.42)
S_{1960}	0.06210	0.05013	0.02123	-0.24233	0.01353
	(1.26)	(0.68)	(0.28)	(1.33)	(0.21)
S_{1970}	0.03732	0.02712	0.02362	-0.23070	-0.00533
	(0.73)	(0.42)	(0.34)	(1.15)	(0.08)
y_{t-1}	0.91608	0.44349	0.11605	0.88046	0.70260
	(41.79)	(5.37)	(1.39)	(4.01)	(7.27)
(year=70)	0.55900	0.24840			0.59849
	(9.12)	(2.17)			(7.52)
(year=80)	0.40168	0.27536	0.36002	0.06115	0.39768
	(5.65)	(3.34)	(3.94)	(0.26)	(2.80)
Constant	0.27154	4.52970	-0.26536	-0.35023	1.63804
	(1.66)	(6.07)	(3.46)	(1.12)	(3.14)
Observ.	292	292	184	184	286
Nr cntries	108	108	102	102	102
SR return[a]	0.15341	0.12648	0.13586	0.07902	0.16034
	(3.41)	(1.64)	(2.29)		
LR return[a]	0.72167	0.14200	0.12947	0.82301	0.41616
	(5.36)	(2.55)	(1.97)		
Sargan				5.34	6.863
p-value				0.255	0.867

Robust t statistics in parentheses. The Arellano-Bond GMM and the Blundell-Bond system GMM estimator assume that the regressors are predetermined (not necessarily exogenous). The coefficient estimates in the table are 1-step estimates since the efficient 2-step estimates are prone to overfitting.

a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Table IX.
GDP: Effect Measurement Error

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	5 year changes		10 year changes		20 year changes		20 year changes Kyriacou data	
S_t	0.06276	0.03991	0.24335	0.08546	0.29273	0.15236	0.24317	0.13828
	(1.03)	(2.74)	(3.86)	(4.11)	(2.73)	(3.33)	(2.98)	(5.07)
S_t^2	-0.00293		-0.00848		-0.01655		-0.00989	
	(1.16)		(2.33)		(2.39)		(1.60)	
S_{1965}	0.09728							
	(1.30)							
S_{1970}	-0.00882		-0.09705					
	(0.16)		(1.97)					
S_{1975}	0.01557							
	(0.22)							
S_{1980}	-0.01051		-0.06732					
	(0.19)		(1.28)					
S_{1985}	0.04885							
	(0.70)							
LR: S	1.57804	0.49367	1.45022	0.48212	0.90067	0.31174	0.79306	0.40663
	(4.85)	(5.41)	(3.29)	(4.29)	(4.21)	(5.54)	(1.64)	(6.91)
LR: S^2	-0.07032		-0.06873		-0.04733		-0.03079	
	(3.68)		(4.61)		(2.39)		(3.92)	
S_{t-1}	0.00931	-0.02249	-0.12164	-0.05574	-0.05754	-0.07883	-0.02842	-0.03305
	(0.15)	(1.54)	(1.92)	(2.62)	(0.48)	(1.63)	(0.34)	(1.27)
S_{t-1}^2	-0.00028		0.00272		0.00419		0.00155	
	(0.11)		(0.72)		(0.46)		(0.19)	
S_{1960}	-0.12359							
	(1.65)							
S_{1965}	-0.01668		0.06210					
	(0.31)		(1.26)					
S_{1970}	-0.03793							
	(0.55)							
S_{1975}	-0.01620		0.03732					
	(0.30)		(0.73)					
S_{1980}	-0.06202							
	(0.91)							
y_{t-1}	0.95433	0.96470	0.91608	0.93837	0.73887	0.76413	0.72920	0.74121
	(85.54)	(103.69)	(41.79)	(45.47)	(14.95)	(14.78)	(10.20)	(10.77)
(year=65)	0.27446	0.15944						
	(5.84)	(7.02)						
(year=70)	0.29379	0.16992	0.55900	0.34489				
	(6.28)	(7.71)	(9.12)	(10.21)				
(year=75)	0.21895	0.11297						
	(4.24)	(5.22)						
(year=80)	0.23575	0.09884	0.40168	0.21204				
	(4.63)	(4.62)	(5.65)	(6.54)				
(year=85)	0.03153	-0.02284						
	(0.62)	(1.08)						
Constant	0.16879	0.24039	0.27154	0.38155	1.85719	1.94997	1.72104	1.87075
	(1.95)	(3.25)	(1.66)	(2.34)	(4.93)	(4.87)	(3.56)	(3.84)
Observations	607	607	292	292	97	97	79	79
R-squared	0.98	0.98	0.95	0.95	0.89	0.88	0.91	0.90
SR return[a]	0.03171	0.03991	0.15341	0.08546	0.11727	0.15236	0.13830	0.13828
	(0.62)	(2.74)	(5.36)	(4.11)	(5.12)	(3.33)	(4.77)	(5.07)
LR return[a]	0.83268	0.49367	0.72167	0.48212	0.39892	0.31174	0.46664	0.40663
	(4.97)	(5.41)	(3.41)	(4.29)	(2.34)	(5.54)	(5.35)	(6.91)

Robust t statistics in parentheses.
a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Table X.
GDP: Subsample Robustness

	(1) Full sample	(2) Without 10 countries with highest growth in education	(3) Without 10 countries with highest growth in GDP	(4) Without 10 countries with highest education level	(5) Without 10 countries with highest GDP	(6) Without 10 countries with lowest education level	(7) Without 10 countries with lowest GDP
S_t	0.24335	0.24390	0.18019	0.21819	0.22457	0.22728	0.23387
	(3.86)	(3.34)	(3.00)	(3.16)	(3.34)	(3.52)	(3.49)
S_t^2	-0.00848	-0.01097	-0.00981	-0.00531	-0.00592	-0.00672	-0.00696
	(2.33)	(2.80)	(2.67)	(1.09)	(1.38)	(1.77)	(1.78)
S_{1970}	-0.09705	-0.07410	-0.00388	-0.10490	-0.10580	-0.09715	-0.11475
	(1.97)	(1.26)	(0.08)	(2.02)	(2.08)	(1.98)	(2.26)
S_{1980}	-0.06732	-0.04650	0.00191	-0.06799	-0.05854	-0.08568	-0.07981
	(1.28)	(0.78)	(0.04)	(1.16)	(1.05)	(1.60)	(1.49)
LR: S	1.45022	1.41359	1.56899	1.16652	1.30400	1.35185	1.27926
	(4.61)	(2.76)	(3.14)	(1.52)	(4.05)	(3.91)	(4.00)
LR: S^2	-0.06873	-0.06502	-0.07191	-0.04383	-0.06390	-0.05971	-0.05542
	(3.29)	(3.95)	(2.36)	(3.50)	(2.76)	(2.63)	(2.69)
S_{t-1}	-0.12164	-0.13788	-0.08758	-0.12305	-0.11593	-0.11709	-0.12212
	(1.92)	(1.93)	(1.50)	(1.79)	(1.78)	(1.84)	(1.84)
S_{t-1}^2	0.00272	0.00609	0.00556	0.00174	0.00060	0.00185	0.00212
	(0.72)	(1.50)	(1.49)	(0.35)	(0.14)	(0.48)	(0.53)
S_{1960}	0.06210	0.04585	-0.02268	0.08450	0.07886	0.06252	0.07833
	(1.26)	(0.78)	(0.47)	(1.58)	(1.52)	(1.29)	(1.56)
S_{1970}	0.03732	0.02607	-0.02017	0.04660	0.03753	0.05053	0.04454
	(0.73)	(0.45)	(0.38)	(0.81)	(0.68)	(0.99)	(0.87)
y_{t-1}	0.91608	0.92500	0.94097	0.91844	0.91669	0.91848	0.91264
	(41.79)	(39.40)	(38.60)	(38.74)	(38.03)	(41.79)	(38.65)
(year=70)	0.55900	0.52383	0.46928	0.51951	0.53614	0.55951	0.58277
	(9.12)	(8.41)	(7.89)	(8.17)	(8.55)	(7.69)	(8.23)
(year=80)	0.40168	0.34373	0.30574	0.37328	0.37098	0.44830	0.44202
	(5.65)	(4.90)	(4.35)	(4.88)	(5.09)	(5.14)	(5.31)
Constant	0.27154	0.24292	0.16569	0.30541	0.29792	0.27488	0.31528
	(1.66)	(1.43)	(0.95)	(1.77)	(1.69)	(1.60)	(1.65)
Observations	292	272	268	267	262	269	266
R-squared	0.95	0.96	0.96	0.95	0.94	0.95	0.94
SR return[a]	0.15341	0.12765	0.07625	0.16187	0.16182	0.15605	0.16012
	(3.41)	(2.42)	(3.45)	(4.58)	(4.54)	(3.45)	(3.47)
LR return[a]	0.72167	0.72436	0.80672	0.70195	0.62666	0.71889	0.69181
	(5.36)	(4.63)	(1.75)	(3.37)	(3.56)	(4.80)	(4.74)
Countries excluded from the sample		Congo Egypt China Hong Kong Jordan Korea Taiwan Austria Bulgaria Germany	Botswana Swaziland Hong Kong Japan Korea Singapore Taiwan Malta Bulgaria Romania	Canada USA Denmark Finland Poland Australia New Zealand Bulgaria Czechoslov Germany	Canada USA Belgium France Germany Italy Netherlands Sweden Switzerland Australia	Benin C. Afr. Rep. Gambia Liberia Mali Mozambique Niger Sudan Nepal P. N. Guinea	C. Afr. Rep. Lesotho Malawi Mali Niger Rwanda Togo Uganda Zaire Myanmar

Robust t statistics in parentheses.
a) The implied return to education is calculated at an average education level of 5.3 years (the sample average in 1990). The standard errors are calculated using the delta method.

Figure I.
Return to Education and Average Education Level Across Countries

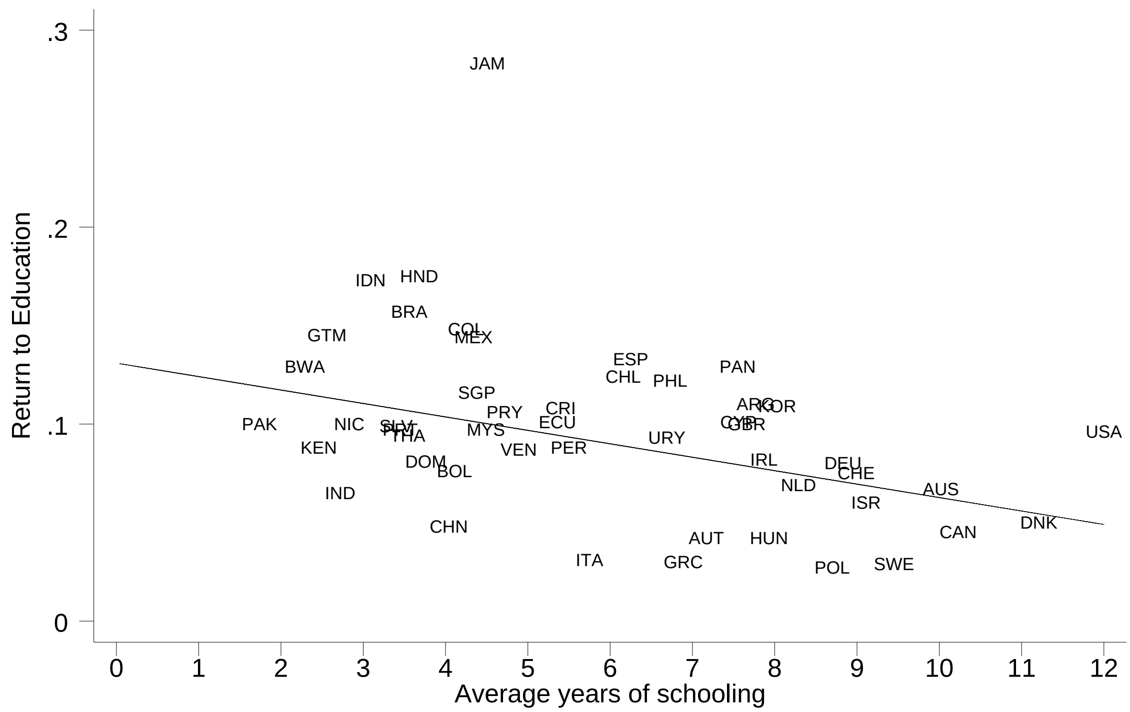

Figure II.
Returns to Education and Inequality Across Countries