

Education, Growth and Income Inequality

Coen Teulings and Thijs van Rens*

First version: January 2001

This version: July 2006

Abstract

Estimates of the effect of education on GDP (the social return) have been hard to reconcile with micro evidence on the private return to schooling. We present a simple explanation combining two ideas: imperfect substitution and endogenous skill-biased technological progress and use cross-country panel data on inequality and GDP to test these ideas. A one-year increase in the level of education reduces the private return by 2 percentage points, consistent with Katz-Murphy's [1992] elasticity of substitution. We find no evidence for reversal of this initial effect as in Acemoglu [2002]. In the short run, the social return equals the private return.

JEL classification: E24, J24, O15

Keywords: growth, inequality, education, private and social return to schooling, compression effect

*Coen Teulings, CPB Netherlands Bureau for Economic Policy Analysis and University of Amsterdam, C.N.Teulings@cpb.nl. Thijs van Rens, CREI and Universitat Pompeu Fabra, thijs.vanrens@upf.edu. We thank Miguel Portela for excellent research assistance and Mikael Lindahl and Daniele Checchi for sharing their datasets. We are also grateful to Roland Bénabou, Richard Blundell, Steve Bond, Maurice Bun, Antonio Ciccone, Rema Hanna, Alan Krueger, Jong-Wha Lee, Mikael Lindahl, Francesc Ortega, Miquel Pellicer, Miguel Portela, Giorgio Primiceri, Robert Shimer, participants of the 2002 annual meeting of the Society of Labor Economists and two anonymous referees for helpful comments. Thijs van Rens gratefully acknowledges financial support from the Generalitat de Catalunya, DURSI, grants Beatriu de Pinós and 2005SGR 0049, and the Spanish Ministry of Education and Sciences, grant SEJ2005-01124.

1 Introduction

Under perfect competition, the social rate of return to education, the effect on log GDP of increasing the education level of a country’s workforce by one year, equals the Mincerian private rate of return. This prediction seems to be contradicted by the data: in growth regressions, the education level seems to have little effect on the *level* of GDP, but a strong effect on its *growth rate*. We present a simple model that reconciles the empirical growth literature with the micro evidence on the private return to schooling. This result depends on two commonly accepted ideas: imperfect substitution between worker types and endogenous skill-biased technological progress. The model suggests an empirical strategy to estimate both the social and the private return from cross-country panel data. Previous papers have used data on GDP to analyze the social return to education. Our idea is to use data on income inequality to estimate the private return as well. We find that the joint evolution of the social and the private return to education is consistent with our model, with previous evidence on the degree of substitution between worker types and with estimates of the private return from micro data. In the short run, the social return equals the private return; in the long run, however, endogenous technological progress creates an externality that increases the social return substantially above the private return.

Imperfect substitution between worker types provides the identifying assumption that allows us to estimate the private return from inequality data. If workers with various levels of education were perfect substitutes, relative wages would be independent of the distribution of human capital. However, if high and low skilled workers are imperfect substitutes [Katz and Murphy 1992], then the rate of return to schooling must be negatively related to the average education level of the workforce: raising the average education level makes low-skilled workers scarcer, which increases their wages. At the same time, it increases the supply of highly educated workers and lowers their relative wages. As a result, the return to human capital falls.

Without externalities in education, the social rate of return to education equals the private rate, so that 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 negative effect on income inequality. We formally model these effects in an assignment model with heterogeneous workers and heterogeneous jobs; see Teulings [1995, 2005]. In this model, highly educated workers have a comparative advantage in complex jobs; the return to education is therefore higher for more complex jobs. When the supply of highly educated workers increases, some of these workers accept less complex jobs, where their human capital has a lower return. The model thus predicts a negative relationship between the aggregate supply of education and its rate of return.

There is a large literature on the effect of education on inequality. Typically, these models focus on two effects: a more unequal education distribution implies more income inequality (the ‘composition’ effect) and a higher average education level implies less income inequality (the ‘compression’ effect) [Knight and Sabot 1983]. Becker and Chiswick [1966] were the first to show that in the data (across regions in the US) the average level of education is indeed negatively correlated with income inequality. Since

then, the same relation has been shown to hold also across countries and within countries over time [De Gregorio and Lee 2002]. In this paper, we focus on the compression effect and derive joint implications for GDP growth and income inequality.

Contrary to the predictions of our model, previous research has documented a relationship between education *levels* and GDP *growth* across countries, but no effect of changes in education on changes in GDP [Benhabib and Spiegel 1994; Barro and Sala-i-Martin 1999]. 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 which there are strong externalities of investments in education. For example, Barro and Sala-i-Martin argue that a higher education level enables the labor force to deal more effectively with technological innovations, yielding a positive relationship between the level of human capital and the growth of output. Empirical studies, however, have failed to find evidence of externalities [Heckman and Klenow 1997; Acemoglu and Angrist 1999].

The static, perfectly competitive, assignment model described above can be extended to allow for dynamics caused by firms' decisions to invest in new technologies and by positive externalities in knowledge accumulation. Following Acemoglu [2002], we argue that investments in new, skill-biased technologies are more profitable when educated workers are more abundant. Since an increase in the average education level reduces the rate of return to education, it makes the application of skill-biased technologies more profitable and induces firms to invest in new technologies. Assuming 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. However, 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 and possibly even reverses the initial negative effect of a higher average education level on its private return.

Causality is always an issue for empirical research in this area: does a higher education level lead to higher GDP or is it the other way around? Indeed, Bils and Klenow [2000] have argued that the 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 We argue that in a dynamic panel setting, time-lags in the causation from GDP to the average schooling level help to solve the endogeneity problem. Consider an exogenous increase in GDP, which causes an increase in the desired education level. To realize the increase in education, the political system first has to decide how much of the additional tax revenues to spend on education. Then, new teachers have to be trained and new educational facilities have to be built. The first cohorts will benefit from improved training only after these changes have been made. And these cohorts of better educated students enter the labor market only several years

¹Bils 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 negative correlation between growth and the return to experience.

later. Therefore, we argue that it is reasonable to assume that GDP affects education with at least a ten-year lag.

We find strong support for a negative relationship between the supply of human capital and its return. Moreover, the compression effect we find in the inequality and in the GDP data is very similar: a one-year increase in the stock of human capital reduces its return by 2 percentage points. This estimate is consistent with Katz and Murphy's [1992] and Ciccone and Peri's [2005] estimates of the elasticity of substitution between low and high skilled workers in the US. Moreover, our estimates for the private return to education from inequality data are also consistent with Mincerian returns to schooling estimated from microdata in several countries. The social return to education approximately equals the private return in the short run, once we control for country-specific fixed effects and endogenous technological progress. This is a considerable step forward for the growth literature, which has typically found that the effect of increases in education on growth is insignificant. The social return to education is much higher in the long run than it is in the short run. 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 suggests the presence of strong positive externalities of education that are realized slowly over time, for instance because a higher schooling level spurs technological development. However, we do not find support for Acemoglu's [2002] prediction that this endogenous technological progress is skill-biased, because we do not observe an increase in the private return in the long run.

Despite the fact that our estimate of the long run social return is lower than estimates found in previous studies, it is still quite large. Based on our estimates, the long run return from investments in education over the sample period is higher even than actual GDP growth. This relates our analysis to O'Neill [1995]. He observes that huge investments in human capital by LDCs have not contributed to a convergence in GDP between LDCs and the industrialized world, and offers an explanation based 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 this explanation: skill-biased technological progress has shifted the terms of trade against developing countries that produce commodities with a low capital intensity. For this reason, countries that did not invest in human capital would have experienced negative productivity growth.

The paper is structured as follows. Section 2 presents 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 structure of the economy

Consider the long run growth path of an economy with physical and human capital, along the lines of Teulings [1995, 2005]. 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,

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

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. For simplicity, we assume that productivity depends only on the difference between the skill level of the worker and the complexity level of the task,

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

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 in order to maximize profits.

$$P_t(c) Y_t(c) - W_t(s) - RK_t(c) \quad (2)$$

²The 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) where all c -type commodities enter directly into the utility function. These interpretations yield exactly the same results.

where $W_t(s)$ is the market wage of an 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,

$$RK_t(c) = \alpha P_t(c) Y_t(c) \quad (3)$$

$$W'_t(s_t(c)) = (1 - \alpha) P_t(c) Y_t(c) h'(s_t(c) - c) \quad (4)$$

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

$$W_t(s_t(c)) = (1 - \alpha) P_t(c) Y_t(c) \quad (5)$$

Using this result, first-order condition (4) becomes,

$$w'_t(s_t(c)) = h'(s_t(c) - c) \quad (6)$$

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,

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

where $\phi(\cdot)$ 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, rises 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 11).

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 the zero-profit condition (5), the 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 task, 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 logarithms, the market clearing condition can be written as,

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

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

$$s_t(c) = c - C_t + S_t \quad (9)$$

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. Aggregate log output per capita is found by substituting (9) into differential equation (8):⁴

$$y_t = h(S_t - C_t) + \log F_t \quad (10)$$

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

⁴In a previous version of this paper [Teulings and Van Rens 2003, section 2.1.2] we show that C_t can be endogenized by letting composing firms choose the average complexity level of inputs; i.e. $\log F_t = f(C_t, t)$. C_t is increasing in S_t due to the comparative advantage of high-skilled workers in more complex jobs: 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, raising the profitability of high C_t technologies. This is consistent with Caselli and Coleman's [2002] finding of a positive correlation between computer use and the average education level of a country's workforce. When the cross-derivative of f is positive, $f_{12} > 0$, the model displays skill-biased technological progress.

As noted earlier, a joint increase in F_t and C_t represents skill-biased technological change, whereas an increase in F_t that does not affect C_t is skill-neutral technological progress.

2.2 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

$$w'_t(s) = w'_t = h'(S_t - C_t) \quad (11)$$

Equation (11) 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 than the variance of the distribution of complexity levels, the Mincer equation is concave, and in the opposite case it is convex [Teulings 2005]. There is a simple intuition for this result. A larger variance of the skill distribution makes skill types around the median scarcer compared to types in the tails of the 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.

The social rate of return to education is obtained by taking the derivative of equation (10) with respect to the average level of schooling,

$$\frac{dy_t}{dS_t} \equiv y'_t = h'(S_t - C_t) = w'_t \quad (12)$$

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 returns to education vary with the average education level of the workforce. The derivative of expression (12) 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.⁵

⁵The 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 (11) 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 from microdata on the linearity of the Mincerian earnings function, we shall interpret our estimation results under this assumption.

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 percentage increase in the value of its stock:

$$\gamma \equiv -\frac{dw'_t/w'_t}{w'_t dS_t} = -\frac{h''}{h'^2} \quad (13)$$

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 there 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 [2005]):

$$\gamma = \frac{1}{\eta_{\text{low-high}} D_t} \quad (14)$$

where D_t denotes the variance of log wages. Using Katz and Murphy's [1992] estimate (from US time series) of $\eta_{\text{low-high}} = 1.4$ or Ciccone and Peri's [2005] estimate (from variation across US states) of 1.5 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 (13) and (14) to compare our estimates for the degree of substitution between worker types with Katz and Murphy's results.

In summary, the basic model yields the following testable implications: the private and the social return to education are equal, and both are 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.

2.3 Endogenous technological progress

In the long run, an increase in the average education level in a country may affect the private and social returns to education in ways not described by our static model. A leading possibility is that the higher education level of the workforce spurs investments in new technologies, which we have so far treated as exogenous. This would be the case if the production of knowledge is more human capital intensive than the production of the consumption good, as in Uzawa [1965], Lucas [1988] and Barro and Sala-i-Martin [1999, chapter 5]. There may also be externalities in knowledge production, so that investments in new technologies by one firm benefit other firms as well. In this section we consider the predictions of an extension to the basic model, in which composing firms can invest in new technologies by allocating a fraction of their workforce to the production of knowledge.⁶

We first discuss the long run effects of a change in the average level of education of the workforce S_t . In steady state, the level of technological development must be such

⁶ A formal discussion can be found in a previous version of this paper [Teulings and Van Rens 2003, section 2.2].

that the cost of an additional unit of investment in knowledge equals the annuity value of revenues from that investment. If there are externalities in knowledge production, then the cost of investing more in knowledge must be equal to the private revenue for the investing firm. The steady state stock of knowledge is increasing in the average level of education in the workforce. The reason is that, conditional on the level of technology, a higher level of education implies a lower return to human capital w' , so that it is less costly to use high skilled workers to develop new technologies.

The long run social return to education consists of two parts: the direct effect of education on output and the indirect effect through the higher steady state level of technology. Unlike in the basic model, the long run social return to education exceeds the private return, even if there are no spillovers in the production of new technologies. This result simply reflects the return on investments in new knowledge. In the presence of knowledge spillovers, the social return exceeds the private return by even more, since every firm benefits from the investment of all other firms.

Contrary to the basic model, the long run effect of an increase in the average level of education on the private return to education is not necessarily negative. The counteracting mechanism is similar to that in Acemoglu [2002]: a greater supply of human capital induces investments in new technology. If these innovations to technology are skill-biased, they raise the private return to human capital. Depending on the degree of skill bias, the increase in the private return because of technological progress may be smaller or larger than the direct decrease in the return because of the increased supply of human capital. The net long run effect of the average education level on the private return to schooling is an empirical question.

Now consider the transition dynamics. Suppose the economy is initially in a steady state equilibrium, when the average level of education S_t permanently increases by one year. Since the current level of technological development is too low compared to the new steady state, firms immediately start investing in knowledge production. And because knowledge production is relatively skill intensive, this raises the demand for human capital and therefore its return. Compared to the basic model, the negative immediate effect of an increase in the supply of human capital on the private return to schooling is partly offset by an increase in the demand for education for the investment in new knowledge. The net effect of both forces is ambiguous, depending on the degree of skill intensity of knowledge production. Notice that the reason for the ambiguity is different for the ambiguity in the sign of the long run response. In the short run, the skill intensity of investment in new knowledge counteracts the standard substitution effect. In the long run, the skill bias of new technologies is the counteracting force.

Like the effect on the private return, the immediate effect of an increase in the level of human capital on the social return to education is also counteracted by increased investments in new technologies. The sign of the net effect on the social return is again ambiguous, but the short run social return is unambiguously smaller than the private return. The reason is that some workers are assigned to knowledge production and are no longer available for current production.⁷

⁷Notice that there is an accounting issue whether or not GDP includes knowledge production. Since

Summarizing, in a dynamic extension to the model with endogenous technological progress, the short run responses to an increase in the average education level in a country of both the private and the social return are ambiguous, but the social rate of return must be smaller than the private return because more workers are allocated to knowledge production, which does not immediately contribute to output. In the subsequent transition path, GDP rises unambiguously, while the private rate of return can either rise or fall, depending on which effect dominates: the increasing level of (skill-biased) technological development or the lower level of investment in knowledge.

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 because our empirical strategy relies crucially on this assumption. We assume that if workers of different education levels are imperfect substitutes, an increase in the supply of human capital will decrease its return. Linearizing equation (11) we get,

$$w'_{jt} = \alpha_1 - \alpha_2 S_{jt} \quad (15)$$

where j indexes countries. We estimate equation (15) using a cross-country sample of private returns to education estimated from microdata by various authors and compiled by Bils and Klenow [2000]. The estimates in table I show a clear negative relationship between the return to education and the average schooling level in a country. The crucial coefficient α_2 is 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 additional year of education. The private rate of return to education is 11% at the sample average education level in 1990 (5.3 years), 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 a measure of the degree of imperfection in the substitution between types of labor, see equation (13):

$$\gamma_{jt} \equiv \frac{dw'_{jt}/dS_{jt}}{w'^2_{jt}} = \frac{\alpha_2}{(\alpha_1 - \alpha_2 S_{jt})^2} \quad (16)$$

For $\alpha_2 S_{jt} < \alpha_1$, the compression elasticity is monotonically increasing in the average education level S_{jt} . The compression elasticity equals 0.60 for the average education level in 1990, and 3.44 for the United States. The latter number is similar to the value of 2 based on Katz and Murphy's [1992] and Ciccone and Peri's [2005] estimates of the elasticity of substitution between high school and college graduates for the US, see

it seems that these investments are largely non-tangible, we assume that at least part of knowledge production is not counted as output.

equation (14). These conclusions do not change if we weight countries by log GDP per worker as in column (2) or eliminate the outlier Jamaica, see column (3).

Because the estimates in table I are based on cross-sectional variation between countries only, they have several limitations. First, since we have only one estimate for the return to schooling in each country, we cannot control for differences across countries. For example, if institutions to protect property rights further the accumulation of human capital and also reduce rent extraction, it is impossible to disentangle these two effects on the return to education from the cross-sectional variation alone. Second, a cross-section does not allow for inferences 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

The accumulation of human capital reduces its return, thereby compressing the wage distribution. But education clearly is not the only factor yielding wage differentials between workers. Therefore, we extend the Mincer equation to allow for other worker characteristics,

$$w_{jt}(s, u) = \alpha_{0jt} + w'_{jt}s + \sigma u \quad (17)$$

where u is a standard normal random variable representing other worker characteristics and σ is their standard deviation. Let V be the variance of years of schooling s , and ρ the correlation between s and other worker characteristics u . Then, the variance of log wages, or wage dispersion, D_{jt} is given by

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

We use equation (15) to substitute for w'_{jt} and get the following expression for wage dispersion,

$$D_{jt} = \theta_0 - \theta_1 S_{jt} + \theta_2 S^2_{jt} \quad (18)$$

where

$$\begin{aligned} \theta_0 &\equiv \alpha_1^2 V + 2\alpha_1 V^{1/2} \sigma \rho + \sigma^2 \\ \theta_1 &\equiv 2\alpha_2 \alpha_1 V + 2\alpha_2 V^{1/2} \sigma \rho \\ \theta_2 &\equiv \alpha_2^2 V \end{aligned}$$

It is important to note that both θ_1 and θ_2 would be zero if $\alpha_2 = 0$, and a proper test of perfect substitution between workers types is to test the joint restriction $\theta_1 = \theta_2 = 0$. The θ parameters depend on the variance of the education distribution V , which we assume to be constant over time and across countries. It is difficult to provide a structural model including the effect of variation in V , because in that case the Mincer equation is no longer linear (see Teulings [2005]). Instead, we adopt a pragmatic approach, adding an additive control term $\theta_3 V_{jt}$ to equation (18). If we assume that capital income is proportional to labor income, so that the log wage distribution and the log income distribution differ only by their first moment, then equation (18) holds for income inequality as well. The equation can then be estimated using panel data on income inequality.

We allow for dynamics in the effect of S_{jt} on D_{jt} and y_{jt} , and estimate,⁸

$$\begin{aligned} D_{jt} &= \theta_{0t} - \theta_1 S_{jt} + \theta_2 S_{jt}^2 + \theta_3 V_{jt} + \bar{\theta}_1 S_{jt-1} - \bar{\theta}_2 S_{jt-1}^2 - \bar{\theta}_3 V_{jt} + \phi D_{jt-1} + u_{jt} \\ y_{jt} &= \beta_{0t} + \beta_1 S_{jt} - \frac{1}{2} \beta_2 S_{jt}^2 - \bar{\beta}_1 S_{jt-1} + \frac{1}{2} \bar{\beta}_2 S_{jt-1}^2 + \psi y_{jt-1} + v_{jt} \end{aligned} \quad (19)$$

where u_{jt} and v_{jt} are error terms. We allow θ_{0t} and β_{0t} to vary over time in order to account for (skill-neutral) technological progress.⁹ The (short run) social return to education,

$$y'_{jt} = \beta_1 - \beta_2 S_{jt} \quad (21)$$

is easily comparable to expression (15) for the private return, $w'_{jt} = \alpha_1 - \alpha_2 S_{jt}$. The short run first-order effect of an increase in S_{jt} on inequality and output is given by θ_1 and β_1 respectively, and the long run effects can be calculated as $(\theta_1 - \bar{\theta}_1) / (1 - \phi)$ and $(\beta_1 - \bar{\beta}_1) / (1 - \psi)$. The short and long run effects are equal if $\bar{\theta}_1 = \phi \theta_1$ for the private return or $\bar{\beta}_1 = \psi \beta_1$ for the social return. If the short and the long run return to education are not equal to each other, that would be evidence in favor of endogenous technological progress as discussed in section 2.3.

Finally, consider the role of physical capital in a regression such as (20). 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, this implies that log aggregate capital differs from log GDP only by a constant:

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

Hence, we cannot 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 $h(S_t - C_t)$ are perfectly collinear. Although this is probably an extreme case, which depends on the specific assumptions about 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 the presence of measurement error in both variables, the relative magnitudes of the coefficients of both factor inputs merely reflect the relative precision of their measurement. Krueger and Lindahl [2001] argue that capital data are correlated to output by construction, since investment data are used to construct both series. Hence, the measurement error in both series is likely to be correlated. This explains why they find α to be much higher than one would expect based on conventional estimates of the capital share in output of about 0.35. Our estimates of (20) including capital as an explanatory variable confirm this finding. Consequently, we omit capital in most of our regressions and apply the reduced-form equation (20) in our preferred specifications.¹⁰

⁸Notice that equation (20) for GDP can be rewritten in error correction form as

$$\Delta y_{jt} = \tilde{\beta}_{0t} + \beta_{1t} \Delta S_{jt} - \frac{1}{2} \beta_{2t} \Delta S_{jt}^2 + (\beta_{1t} - \bar{\beta}_{1t}) S_{jt-1} - \frac{1}{2} (\beta_{2t} - \bar{\beta}_{2t}) S_{jt-1}^2 - (1 - \psi) y_{jt-1} + v_{jt}$$

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

⁹We also estimated versions of (19) and (20), in which θ_1 , $\bar{\theta}_1$, β_1 , and $\bar{\beta}_1$ are allowed to be time-varying, in order to account for (exogenous) skill-biased technological progress.

¹⁰Alternatively, we could fix the coefficient for capital to some value, as Krueger and Lindahl [2001]

3.3 Identification and estimation

A serious problem in the estimation of equation (20) for GDP and (19) for wage dispersion is reverse causation: does a rise in education cause growth in GDP, or does growth cause rising educational attainment? Our approach to this problem is to use the dynamic structure of the effects. In particular, we assume that contemporaneous changes in GDP do not affect the average education level of the workforce in that same observation period. We follow Krueger and Lindahl [2001] and use a ten-year timeframe to alleviate the measurement error bias in the coefficient on ΔS_{jt} . To interpret our estimates, we therefore need to assume that a change in GDP takes at least ten years to affect the average education level in a country. There are several reasons to believe this is a realistic assumption. It takes time before an increase in GDP leads to an increase in the budget of the education system. Then, it takes time to actually build new schools and train new teachers. Finally, the new generation of students that benefits from the increased expenditures on schooling takes time to finish school and enter the labor market.

Another problem is the possibility that there are time-varying third factors that affect both the level of GDP and human capital formation, e.g. better institutions for the protection of property rights. If these factors have an immediate effect on human capital formation and a delayed effect on GDP some ten years later, then we would conclude incorrectly that the increase in education was responsible for the rise in GDP. Although we cannot rule out this possibility, we are not aware of time-varying variables that have been shown to have a strong impact on GDP. The quality of institutions tends to be quite persistent. The particular time pattern that is needed for these variables to cause bias in our estimates makes it even less likely that this is a serious concern.

Estimating (19) and (20) 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 (20) 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

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

Clearly, OLS is inconsistent in this case, because $y_{j,t-1}$ is correlated with f_j . First differencing the equation eliminates the fixed effect, but now the component $\varepsilon_{j,t-1}$ in $\Delta\varepsilon_{jt}$ is correlated with $\Delta y_{j,t-1}$ (and possibly also with ΔS_{jt} via the reverse causation equation).

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 preferable 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.

A consistent and efficient GMM estimator uses the following moment conditions:

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

These moment conditions give rise to the dynamic panel data estimator set out in Arellano and Bond [1991] and applied to the growth literature starting from Caselli et al. [1996].¹¹ Because the regressors in this paper include not only education but also education squared, we include lags of both education and education squared in the instrument set.

Since we use ten year time intervals, the time dimension of our panel is very short (at most five periods), so that our estimates using this estimator are imprecise. 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:

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

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

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

These additional moment conditions give rise to the Blundell-Bond system estimator.¹² Because output is highly persistent, the instruments in the Arellano-Bond estimator are weak, so that the estimator is biased in finite samples (Blundell and Bond [1998], Hahn, Hausman and Kuersteiner [2004]). Bond et al. [2001] reestimate the dynamic growth regressions in Caselli et al. and show that in the cross-country growth context, the estimator is indeed poorly behaved. For this reason, they advocate the use of the Blundell-Bond system estimator for panel growth regressions.

3.4 Data sources

We draw data from four sources: the Barro and Lee [2000, 1996, 1993] data on educational attainment, the World Income Inequality Database [UNU-WIDER 2005], the Penn World Table [Heston, Summers and Aten 2002] for GDP per worker and the data on capital constructed by Klenow and Rodriguez-Clare [2005]. The last update of the Barro and Lee dataset contains detailed data on educational attainment for 142 countries for the period 1960-2000 in intervals of five years. Barro and Lee report the fraction

¹¹Islam [1995] also argues in favor of controlling for fixed effects in growth regressions. However, the estimator used in that paper assumes strict exogeneity (rather than predeterminedness) of the regressors.

¹²We implement this estimator using Roodman's [2005] `xtabond2` module for Stata. For earlier versions of the paper we also used the Ox package by Doornik et al. [2002] and Bond's DPD software for Gauss.

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.¹³

Measurement error is a serious problem in all data on educational attainment, including the Barro and Lee dataset (see Krueger and Lindahl [2001]). De la Fuente and Doménech [2001] improved the data for 21 OECD countries by using previously unexplored sources and by removing sharp breaks in the data. These authors avoid the use of enrollment data. But their data have two drawbacks: the small number of countries and the rather ad hoc choices the authors have to make when correcting unexplainable breaks in time series.¹⁴ Cohen and Soto [2001] extend De la Fuente and Doménech’s work to several other countries (contrary to De la Fuente and Doménech however, they use enrollment data when needed). They construct education data for 95 countries, trying to “minimize the extrapolations and keep the data as close as possible to those directly available from national censuses.” Portela, Alessie, and Teulings [2004] address the adequacy of the perpetual inventory method in the Barro and Lee data directly. They find clear evidence for a large and systematic bias in the data points constructed by the perpetual inventory method and develop a statistical procedure to correct for this bias. However, they also show that for the same growth regressions as used in this paper, using original or corrected data makes little difference. In light of this conclusion, we use the original Barro and Lee data.

The World Income Inequality Database (WIID) data on income inequality were taken from a large number of studies and assessed on their comparability and the quality of the estimates. We use the new quality label provided in version 2.0a of the WIID, which combines and improves the quality ratings in the Deininger and Squire [1996] data with older versions of the WIID. We use only data for which the underlying concepts of income and income receiving units are known, and for which the quality of the concepts and the survey is judged to be sufficient in terms of coverage, questionnaires and data collection.

¹³Barro 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:

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

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 [2004]:

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

¹⁴De la Fuente and Doménech are aware of these problems and admit that “the construction of our series involves a fair amount of guesswork”.

For this reason, the number of observations included for some countries is smaller than in the Deininger and Squire [1996] data, even though the time period covered is much longer. We thus address Atkinson and Brandolini’s [2001] criticisms about the quality and consistency of the earlier datasets as well as we can. We discuss the Atkinson-Brandolini critique in more detail when we present our estimation results in section 3.5. When the inequality data contain missing values, we construct data for ten year intervals from 1960 to 2000 by linear inter- and extrapolation with a maximum span of two years from an actual datapoint.¹⁵ We calculated the variance of log income from the Gini coefficients, assuming that log income is distributed normally (see appendix A). Table II provides summary statistics for the variables in the combined dataset.¹⁶

3.5 Estimates of the private return from inequality data

Table III presents estimates of equation (19). Since the data on income inequality are not fully comparable across countries and years, we control for differences in the way the Gini is calculated. In particular, we include dummy variables to control for the unit of analysis in the underlying microdata (household or person), the income definition used to calculate the Gini coefficient (gross income, disposable income or consumption), the income share unit (family or household) and the equivalence scale used vary across different observations.

In columns (1) through (4) we present estimates of the model without controlling for fixed country effects. Column (2) adds the variance of education, column (3) introduces the second-order effect of education and column (4) weighs observations by log GDP per worker. In all these regressions, the education variables are jointly insignificant. Moreover, in the presence of fixed effects, the estimate of the long run return is biased upward because correlation of the lagged dependent variable D_{t-1} with the fixed effect biases its coefficient towards one.

In columns (5), (6) and (7) we control for country-specific fixed effects, identifying the estimates only off the variation in inequality within countries over time. In these regressions, the education variables are jointly significant and all have the sign expected on the basis of our model. The within-group estimator in column (5) is inconsistent and underestimates the coefficient on D_{t-1} .¹⁷ Column (6) presents the consistent and efficient system GMM estimator proposed by Blundell and Bond [1998]. The Sargan/Hansen test cannot reject the overidentifying restrictions of the system estimator (p-value is 100%) and the Arellano-Bond tests detect first-order autocorrelation in the error terms (p-value is 7%) but do not find evidence for higher order autocorrelation (p-value 37% for second-order), as we would expect if the system estimator is correctly specified.

¹⁵We only interpolate between observations that were calculated from the same source by the same definitions. In that case, we use $\hat{x}_t = \frac{p}{p+q}x_{t-q} + \frac{q}{p+q}x_{t+p}$, 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 nearest observation (at most two years apart). This procedure is efficient if the Gini follows a random walk, as is almost true empirically.

¹⁶The complete dataset as well as the Stata codes used to create it are available from the authors.

¹⁷In addition, the standard errors are biased downward because taking deviations from country means introduces autocorrelation in the error term.

In column (7) we explore whether our estimates are sensitive to poor comparability of the inequality data within countries. As pointed out by Atkinson and Brandolini [2001], additive dummy variables may be insufficient to control for changes in definitions of the Gini coefficient. We therefore reestimated the model, including separate dummy variables for all observations with a definitional change. Thanks to improvements in the WIID data and because we constructed the data so as to minimize the number of definitional changes within a country, the estimates are not very sensitive to this robustness check. We also tested whether the results are driven by outliers by excluding various groups of countries from the sample and found the estimates to be robust to this exercise as well.¹⁸

For all specifications that include both education and education squared, we calculate the implied short and long run (private) return to schooling at the sample average of 6.3 years of education. The calculation of the return to education requires values for the variance of the education distribution V , the correlation between education and other worker characteristics ρ , and the variance of those characteristics σ^2 , see equation (18). An estimate for V can be found in table II: $V \cong 19.3$ in 2000. Since we do not have reliable estimates for ρ and σ^2 , the subsequent calculations are based on $\rho = 0$.¹⁹ Despite the small number of observations, the estimates of the short run return are significant and robust across specifications. The estimates of the long run return are less reliable, in part because the coefficient estimates for the lagged education variables are not as precise, but mostly because the long run return is very sensitive to the coefficient on lagged inequality, which is imprecisely estimated and varies across specifications. However, none of the models reveal significant differences between the long and the short run private return, with p-values never below 36% and much higher for the specifications that control for fixed effects. Hence, the data do not support Acemoglu's [2002] 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.

We take the estimates in column (6), which are consistent in the presence of country-specific fixed effects on inequality, as our benchmark. By equation (19), the estimates imply $\alpha_1 = 0.27$ and $\alpha_2 = 0.020$. These parameter values imply a private return at

¹⁸These results are omitted here for brevity but may be found in an earlier version of this paper [Teulings and Van Rens 2003].

¹⁹Ruling out $\rho < 0$, this provides a lower bound on the effect of education on wage dispersion

$$\theta_1 = 2\alpha_1\alpha_2V + 2\alpha_2V^{1/2}\sigma\rho = 2\alpha_2V\left(\alpha_1 + V^{-1/2}\sigma\rho\right) \geq 2\alpha_1\alpha_2V$$

An upper bound can be found by setting $\rho = 1$ and σ^2 equal to the total variance of log wages: $\sigma = D_t^{1/2} \simeq 0.62$ from table II. In that case $V^{-1/2}\sigma\rho \simeq 0.16$, a bit smaller than the implied value for α_1 of 0.27. Hence, setting $\rho = 0$ will not greatly affect the conclusions.

the average education level of 6.3 years in 2000 of 14% and a compression elasticity of $\gamma = 0.96$ by equations (15) and (16). For the US, where the average education level is 12 years, the implied private return is 3%. This estimate is unreasonably low most likely due to the fact that the US has an extremely high education level compared to other countries in the sample. This makes the estimated return less reliable, as it depends more heavily on the estimate of the higher order terms.

Because we find no 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 III are comparable to the cross-section evidence presented in table I. The implied returns to education in table III are somewhat lower than those in table I for countries with high levels of education and somewhat higher for countries with low education levels, but both estimates yield comparable numbers. If there is skill-biased technological change over the sample period, the coefficients θ_1 and $\bar{\theta}_1$ in (19) are not constant over time. We allowed for time-variation in the coefficients on education and education lagged, but found that this makes little difference for the estimates.²⁰

The effect of the variance of the education distribution on income inequality is insignificant and small: a one standard deviation increase in V (6 years² in 2000) decreases income inequality by 0.0138, less than 0.06 of a standard deviation. This suggests that the composition effect of a reduced variance of schooling on the income distribution is less important than the compression effect due to imperfect substitutability of types of labor, which is the main focus of this paper.²¹ However, since we only have a rough proxy for V_t , the low coefficient might also be due to attenuation bias.

Our analysis of the private rate of return to education based on inequality data leads to the following conclusions: (*i*) once country-specific fixed effects are taken into account, 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 consistent with evidence from microdata in different countries, and (*iv*) the long run effect of an increase in the average education level on the private return approximately equals the short run effect and is negative, contradicting Acemoglu's [2002] hypothesis of overshooting.

3.6 Estimates of the social return from GDP data

Estimation results for equation (20) are reported in table IV. Column (1) replicates Krueger and Lindahl [2001, table 3]. The results differ slightly because of our longer time series. The short run social return to education of 8% is roughly equal to our estimate of the private return. The long run return is about seven times higher than the short run effect and exceeds by far any estimate of the Mincerian rate. However, the long

²⁰The implied short run return to schooling goes down from 9.4% to 7.9% for 2000 and the long run return increases from 8.8% to 9.8%. The difference between the two is insignificant and the interaction terms are all individually as well as jointly insignificant (p-value 94%).

²¹In 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 [2005]). Whether the direct composition effect or the indirect general equilibrium effect dominates depends on the parameters of the model.

run effect takes a very long time to materialize because the coefficient of lagged GDP is close to unity. Column (2) adds the 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 a factor of about seven. These results are virtually unaltered when observations are weighted by log GDP per worker, as in column (3). Column (4) adds capital per worker as an explanatory variable. The coefficient for capital per worker is larger than conventional estimates for the capital share of production and the education variables become insignificant, individually as well as jointly. The results confirm the analysis in section 3.2 that the effects of human and physical capital on GDP are hard to disentangle. Given our conclusion that omitting physical capital does not bias the estimates of the social rate of return, we do not include the physical capital in the rest of our regressions.

Columns (5), (6) and (7) control for country-specific fixed effects. Column (5) reports the simple OLS estimator, which is inconsistent and underestimates the coefficient on the lagged dependent variable; and columns (6) and (7) present system GMM estimates. The Sargan/Hansen test borderline rejects the overidentifying restrictions in column (6) at the 3% level. In column (7), we add cross effects of time and education, taking 2000 as a point of reference for the return to education. These cross effects capture skill-biased technological progress and Tinbergen’s [1975] race between education and technology. As time goes by, and keeping constant the average level of education of the workforce, the return to education rises. However, the steady increase in the average level of education over the postwar period in most countries has offset this upward trend in the return to education, as can be seen from the negative second-order effect of education on GDP. The strong effect of skill-biased technological progress can be seen by comparing the first-order coefficients for education in columns (6) and (7). In column (6), that coefficient measures the average return to education at $S_t = 0$ over the whole estimation period. In column (7), the same coefficient is the return in the year 2000. The estimate in column (7) is 18 percentage points higher than that in column (6). The point estimates for the interaction terms indicate that, holding education levels constant, the global return to education would have risen by about 1.7 percentage points in the 1970s, by 2.1 percentage points in the 80s and would have remained constant in the 90s. The interaction terms are jointly significant (p-value 0.02), and adding these interaction effects leads to the Sargan/Hansen test of overidentifying moment restrictions no longer rejecting the model (p-value 63%). For both columns (6) and (7), there is first-order, but not higher order autocorrelation, as expected if the system estimator is correctly specified.²²

We first discuss the implications of our estimates for the short run social return. In this discussion, we take the estimates in column (6) as a point of reference for comparison with the results on inequality.²³ The estimate for the short run social return to

²²The p-values for these tests are given in the caption of table IV.

²³A concern in comparing the estimates in table IV with those in table III is that the estimation samples are not the same because for many countries reliable inequality data are not available. The summary statistics in the last column of table II confirm that countries with higher GDP and education levels are overrepresented in the inequality sample. However, because we use a non-linear specification

education, $\beta_1 = 0.28$, is very close to the average private return, $\alpha_1 = 0.27$. For countries with high education levels, the short run social return is a bit lower than the private return, with $\beta_2 = 0.031 > \alpha_2 = 0.020$ although the difference is not significant.²⁴ At the average education level in our sample in the year 2000, these estimates imply a social return to education of 8%. The results show clear evidence for the second-order effect of education on GDP, which is significant at the 5% level, confirming the hypothesis of imperfect substitution between workers of different education levels. The implication is that investments in a better educated workforce have a much stronger effect on GDP growth in low S_t countries. If high S_t countries grow faster than countries with lower education levels, it must be because of skill bias in technological progress, which shifts technological progress to their advantage (cf. O’Neill [1995]).

Turning to the long run effects, it becomes clear why it is important to control for exogenous skill-biased technological progress. The rate of convergence implied by the estimates in column (6) is 0.6% per year, but increases to the much more reasonable value of 3.6% per year in column (7). The latter estimate is consistent with the 2 to 4% annual convergence found by Bond et al. [2001].²⁵ Skill-biased technological progress implies divergence, as it allows countries with high education levels to grow faster, and failing to control for this effect biases the rate of convergence downward. The implied long run social return, evaluated at the average education level of 6.3 years in 2000, is about 42%, much higher than the private return. This is suggestive of endogenous technological progress as described in section 2.3. However, it takes several decades before a substantial part of the surplus of the long run over the short run return is actually realized. The point estimate for the long run return should be interpreted with care because it is very sensitive to the rate of convergence. Bond et al. [2001] note that there is “a great deal of uncertainty in measuring convergence rates” (p.22). Moreover, the long run effects are more prone to possible reverse causation.

How do our estimates for the social return fit into the global picture of post-war productivity growth? From 1960 to 2000, the average education level of the countries in our sample increased by about 0.069 years per annum. At a long run social return of about 42%, that implies a 2.9% GDP growth per year. Actual growth is only 2% per year over the sample period (see table II). Increased education levels alone seem to account for *more* growth than is actually observed in the data. There are two explanations for this discrepancy. First, 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 over the sample period. Second, skill-biased technological progress shifts the terms of trade at the global level. This hurts countries that do not increase the

for the effect of education on GDP our model captures the differences in the return across the two samples so that the results are not affected. Re-estimating table IV on the sample of countries, for which inequality data are available, substantially deteriorates significance but leaves most of the point estimates and all of the conclusions unaltered.

²⁴Notice that the coefficient on education squared in table IV is $\frac{1}{2}\beta_2$, see equation (20).

²⁵Caselli et.al. [1996] find a much higher convergence rate of about 10% per year. Temple [1999] reports panel estimates from other studies ranging from 0 to as high as 30% per year. However, Bond et.al. [2001] show that these very high convergence rates are driven by weak instrument bias.

education level of their work force. This explanation is confirmed by our estimates: the coefficients on the time dummies show that a country that started in 1960 with an average education level of zero and did not raise its education level since then, would have experienced a *decrease* in GDP from 1970 to 2000 of about 27% or 0.9% per year.

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 equals the private return, and (iii) the long run social return is much higher than the private return, suggesting the importance of endogenous technological progress.

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 a 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 also induces investment in new knowledge, driving up the long run social rate of return to education above the long run private rate. We derived easy-to-interpret relationships between educational attainment, GDP and the social rate of return, and between educational attainment, income inequality and the private rate of return. Our empirical results provide strong support for a negative relationship 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 about 2 percentage points. The estimate for the private return is in line with conventional estimates of the elasticity of substitution between low and high skilled workers [Katz and Murphy 1992, Ciccone and Peri 2005].

The short run social return to education approximately equals the private return. The long run social return, although imprecisely estimated, is clearly much higher than the private return. Our estimates of the GDP equation represent a substantial improvement over the existing growth literature, and we explain why previous studies did not find an effect of increases in education on GDP growth. Partly, this is because of measurement error, as argued by Krueger and Lindahl [2001]. But allowing for dynamics is at least equally important. Krueger and Lindahl find that the estimated social return increases with the time intervals used. Although they attribute this effect to measurement error, it could also be driven by the fact that the long run social return is substantially higher than the short run return.

We find some evidence for exogenous skill-biased technological change. Mainly, however, the very high estimates for the long run social return to education suggest there was enormous *endogenous* technological progress. This endogenous technological progress appears to have been largely skill-neutral and 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. Empirically, however, we do not find support for this implication: a larger supply of human capital reduces the private return to education unambiguously.

A Gini coefficient and the variance of log income

Let $W \in [\underline{W}, \overline{W}]$ 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,

$$Z(W) = \frac{1}{M} \int_{\underline{W}}^W xf(x)dx \quad (22)$$

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 Lorenz 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 be 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*, 40(1), pp.7-72.
- [2] Acemoglu, Daron and Joshua Angrist (1999). How large are Social Returns to Education? Evidence from Compulsory Schooling Laws. NBER WP 7444.
- [3] Acemoglu, Daron, Simon Johnson and James A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(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*, 58(2), pp.277-297.
- [5] Atkinson Anthony B. and Andrea Brandolini (2001). Promise and Pitfalls in the Use of “Secondary” Data-Sets: Income Inequality in OECD Countries As a Case Study. *Journal of Economic Literature*, 39(3), pp.771-799.
- [6] Barro, Robert J. and Xavier Sala-i-Martin (1999). *Economic Growth*. Cambridge: 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*, 32(3), pp.363-394 (data: <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*, 86(2), Papers and Proceedings, pp.218-223 (data: <http://www.nber.org/data> or <http://www.worldbank.org/research/growth/ddbarle2.htm>).
- [9] Barro, Robert J. and Jong Wha Lee (2000). International Data on Educational Attainment: Updates and Implications, CID WP 42.
- [10] Becker, Gary S. and Barry R. Chiswick (1966). Education and the Distribution of Earnings. *American Economic Review*, 56(1/2), Papers and Proceedings, pp.358-369.
- [11] Benhabib, Jess and Mark M. Spiegel (1994). The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data. *Journal of Monetary Economics*, 34(2), pp.143-174.
- [12] Bills, Mark and Peter J. Klenow (2000) Does Schooling Cause Growth? *American Economic Review*, 90(5), pp. 1160-1183.
- [13] Blundell, Richard and Stephen Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87(1), pp.115-143.

- [14] Bond, Stephen (2002). Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. CEMMAP WP CWO09/02.
- [15] Bond, Stephen, Anke Hoeffler and Jonathan Temple (2001). GMM Estimation of Empirical Growth Models. CEPR DP 3048.
- [16] 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.
- [17] Caselli, Francesco, Gerardo Esquivel and Fernando Lefort (1996). Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth*, 1(3), pp.363-389.
- [18] Caselli, Francesco and Wilbur John Coleman (2002). The U.S. Technology Frontier. *American Economic Review*, 92(2), pp.148-152.
- [19] Checchi, Daniele (2004). Does Educational Achievement Help Explain Income Inequality? *Inequality, growth, and poverty in an era of liberalization and globalization*, pp. 81-111, UNU-WIDER Studies in Development Economics. Oxford and New York: Oxford University Press.
- [20] Checchi, Daniele and Luca Flabbi (1999). Income and Educational Distribution Dataset: Various Countries in Panel Format (data: http://www.ecodip.unimi.it/pag_pers/chechchi/chechchi1.htm)
- [21] Ciccone, Antonio and Giovanni Peri (2005). Long-Run Substitutability between More and Less Educated Workers: Evidence from U.S. States 1950-1990. *Review of Economics and Statistics*, 87(4), pp. 652-663.
- [22] Cohen, Daniel and Marcel Soto (2001). Growth and human capital: Good data, good results. OECD Development Centre Technical Paper 179.
- [23] De Gregorio, Jose and Jong-Wha Lee (2002). Education and Income Inequality: New Evidence from Cross-Country Data. *Review of Income and Wealth*, 48(3), pp.395-416.
- [24] Deininger, Klaus and Lyn Squire (1996). A New Data Set Measuring Income Inequality. *World Bank Economic Review*, 10(3), pp.565-591 (data: <http://www.worldbank.org/research/growth/dddeisqu.htm>).
- [25] De la Fuente, Angel and Rafael Doménech (2001). Schooling Data, Technological Diffusion, and the Neoclassical Model. *American Economic Review*, 91(2), pp.323-327.
- [26] Doornik, Jurgen A., Manuel Arellano and Stephen Bond (2002). Panel Data Estimation Using DPD for Ox. <http://www.doornik.com>.

- [27] Hahn, Jinyong, Jerry Hausman and Guido Kuersteiner (2004). Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects. *Econometrics Journal*, 7(1), pp.272-306.
- [28] Heckman, James J. and Peter J. Klenow (1997). Human Capital Policy. mimeo, University of Chicago. <http://www.klenow.com>.
- [29] Heston, Alan, Robert Summers and Bettina Aten (2002). Penn World Table version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002
- [30] Islam, Nazrul (1995). Growth Empirics: A Panel Data Approach. *Quarterly Journal of Economics*, 110(4), pp.1127-1170.
- [31] Katz, Lawrence F. and Kevin M. Murphy (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107(1), pp.35-78.
- [32] Klenow, Peter J. and Andres Rodriguez-Clare (2005). Externalities and Growth. Chapter 11 of *Handbook of Economic Growth*, volume 1A (P. Aghion and S. Durlauf, eds.), pp.817-861.
- [33] Knight, John B. and Richard H. Sabot (1983). Educational Expansion and the Kuznets Effect. *American Economic Review*, 73(5), pp.1132-1136.
- [34] Krueger, Alan B. and Mikael Lindahl (2001). Education for Growth: Why and For Whom? *Journal of Economic Literature*, 39(4), pp.1101-1136.
- [35] 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 WP RR 9126 (dataset available on request).
- [36] Lucas, Robert. E (1988). On the Mechanisms of Development Planning. *Journal of Monetary Economics*, 22(1), pp.3-42.
- [37] O'Neill, Donald (1995). Education and Income Growth: Implications for Cross-Country Inequality. *Journal of Political Economy*, 103(6), pp.1289-1301.
- [38] Portela, Miguel, Rob Alessie, and Coen N. Teulings (2004). Measurement Error in Education and Growth Regression, Tinbergen Institute DP 04-040/3.
- [39] Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update. *World Development*, 22(9), pp.1325-1343.
- [40] Roodman, David (2005). xtabond2: Stata module to extend xtabond dynamic panel data estimator. Center for Global Development, Washington. <http://econpapers.repec.org/software/bocbocode/s435901.htm>

- [41] Summers, Robert and Alan Heston (1991). The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988. *Quarterly Journal of Economics*, 106(2), pp. 327-368 (mark 5.6a: <http://pwt.econ.upenn.edu> or <http://www.nber.org/data>).
- [42] Temple, Jonathan (1999). The New Growth Evidence. *Journal of Economic Literature*, 37(1), pp.112-156.
- [43] Teulings, Coen N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy*, 103(2), pp.280-315.
- [44] Teulings, Coen N. (2005). Comparative Advantage, Relative Wages, and the Accumulation of Human Capital. *Journal of Political Economy*, 113(2), pp.425-461.
- [45] Teulings, Coen N. and Thijs van Rens (2003). Education, Growth and Income Inequality. CEPR DP 3863 or Tinbergen Institute DP 02-001/3.
- [46] Tinbergen, Jan (1975). *Income Distribution: Analysis and Policies*. Amsterdam: North Holland.
- [47] UNU-WIDER (2005). *World Income Inequality Database (WIID)*, version 2.0a. United Nations University - World Institute for Development Economics Research, June 2005. <http://www.wider.unu.edu/wiid/wiid.htm>
- [48] Uzawa, Hirofumi (1965). Optimal Technical Change in an Aggregative Model of Economic Growth. *International Economic Review*, 6, pp.18-31.
- [49] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1), pp.25-51.

Table I
Direct Estimates of Diminishing Returns to Schooling

	(1) OLS	(2) WLS	(3) OLS excl. Jamaica
S_t	-0.00708 (3.23)	-0.00721 (3.41)	-0.00638 (3.68)
Constant	0.15663 (10.33)	0.15725 (10.50)	0.14513 (11.95)
Observations	49	49	48
R-squared	0.36	0.37	0.40

Dependent variable is the return to education estimated from microdata. All columns include time dummies with 1990 as the reference year (the dummy for 1975 was dropped because there is only one observation in that year). WLS regression is weighted by log GDP per worker. Robust t-statistics in parentheses. Return to education data from Bils and Klenow (1998). Education data from Barro and Lee. 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
Summary Statistics

		1960	1970	1980	1990	2000	Total	Est smpl*	Description and source
y_t	mean	8.5694	8.8370	9.1468	9.1877	9.3685	9.0264	10.3219	Log real GDP per worker, 1985 intl prices, chain index (Penn World Table)
	sd	0.9639	1.2009	1.0660	1.0804	1.0987	1.1179	0.6358	
	N	126	135	146	131	130	668	54	
Δy_t	mean		0.0332	0.0274	0.0073	0.0112	0.0203	0.0219	10 year changes in real GDP per worker (annualized).
	sd		0.0236	0.0848	0.0260	0.0225	0.0499	0.0137	
	N		126	135	125	109	495	53	
D_t	mean	0.3255	0.3557	0.3511	0.3676	0.4496	0.3875	0.3779	Variance of log income. Calculated from Gini coefficient (WIID).
	sd	0.1670	0.1681	0.2026	0.2173	0.2372	0.2152	0.2038	
	N	7	11	23	31	36	108	55	
ΔD_t	mean		-0.0036	-0.0064	0.0019	0.0055	0.0013	0.0017	10 year changes in variance log income (annualized).
	sd		0.0053	0.0083	0.0098	0.0094	0.0099	0.0097	
	N		5	11	21	23	60	55	
S_t	mean	3.5256	3.9218	4.7045	5.8108	6.2942	4.8843	8.8195	Average years of education attained by population over 25 years old (Barro and Lee).
	sd	2.5502	2.7302	2.8706	3.0222	2.9350	3.0143	1.9696	
	N	108	111	117	126	112	574	55	
ΔS_t	mean		0.0450	0.0849	0.0880	0.0729	0.0729	0.0806	10 year changes in average years of education (annualized).
	sd		0.0604	0.0658	0.0597	0.0501	0.0614	0.0618	
	N		107	110	111	112	440	55	
V_t	mean	10.4796	11.9090	14.7051	17.4605	19.2567	14.8623	18.3032	Variance of the education distribution (constructed from Barro and Lee data).
	sd	5.5679	5.8179	5.9274	6.0084	6.0673	6.7067	4.8816	
	N	108	111	117	126	112	574	55	
ΔV_t	mean		0.1565	0.2959	0.2838	0.1683	0.2265	0.1445	10 year changes in variance education (annualized).
	sd		0.2809	0.3089	0.3205	0.1996	0.2875	0.3200	
	N		107	110	111	112	440	55	

* Because inequality data are only available for a (non-random) subsample of countries, the last column presents summary statistics for these countries only. This is the estimation sample in table III.

Table III
Income Inequality and the Private Return

	(1) OLS	(2) OLS	(3) OLS	(4) WLS	(5) OLS-FE	(6) Blundell Bond (BB)	(7) BB, incl dummies definition changes
S_t	-0.01184 (0.43)	-0.02593 (1.07)	-0.12954 (1.29)	-0.13344 (1.27)	-0.23449 (1.85)	-0.20326 (3.06)	-0.12625 (2.60)
S_t^2			0.00631 (1.02)	0.00668 (1.04)	0.01176 (1.34)	0.00968 (2.21)	0.00522 (1.74)
V_t		-0.00875 (1.53)	-0.00536 (0.77)	-0.00498 (0.69)	0.00514 (0.59)	-0.00230 (0.38)	-0.00396 (0.90)
S_{t-1}	0.01351 (0.54)	0.02207 (0.92)	0.06604 (0.81)	0.06671 (0.78)	0.05309 (0.93)	0.14499 (3.08)	0.10316 (3.09)
S_{t-1}^2			-0.00293 (0.56)	-0.00311 (0.58)	-0.00131 (0.36)	-0.00737 (2.25)	-0.00466 (2.24)
V_{t-1}		0.00364 (0.69)	0.00212 (0.36)	0.00174 (0.28)	0.00697 (1.61)	0.00088 (0.13)	0.00162 (0.59)
D_{t-1}	0.83183 (5.88)	0.87527 (6.46)	0.84497 (5.28)	0.83861 (5.08)	0.37394 (1.62)	0.63663 (3.35)	0.95507 (11.64)
Observations	55	55	55	54	55	55	55
Nr Countries	24	24	24	24	24	24	24
	<i>Test for joint significance of education and education squared and their lags</i>						
F-statistic	0.22	0.68	1.07	0.90	3.15	13.16	6.68
p-value	0.80	0.51	0.39	0.47	0.06	0.00	0.00
	<i>Implied private return to education</i>						
Short run			0.07187 (2.24)	0.06870 (2.11)	0.09085 2.59	0.09428 (5.56)	0.09536 (4.77)
Long run			0.20215 (1.49)	0.19942 (1.46)	0.10005 2.31	0.08828 (2.15)	0.41933 (0.81)
	<i>Test that long run return equals short run return</i>						
F-statistic			0.87	0.87	0.05	0.02	0.41
p-value			0.36	0.36	0.82	0.88	0.53

Robust t-statistics in parenthesis. All columns include time dummies. Columns (1) through (6) also include dummies for the definition of the Gini coefficient, both contemporaneous and lagged. Column (7) includes dummies for every change in definition of the Gini coefficient within a country (20 in total). The implied return to education is calculated at an average education level of 6.3 years (the sample average in 2000). The standard errors were calculated using the delta method.

The Blundell-Bond system GMM estimator in columns (6) and (7) are 1-step estimates and assume the regressors are predetermined, not necessarily exogenous. The Hansen test of overidentifying restrictions has a p-value of 1.00 in both columns. The Arellano-Bond tests for autocorrelation have the following p-values. Column (6): 0.07 for AR(1), 0.37 for AR(2). Column (7): 0.07 for AR(1), 0.18 for AR(2).

Table IV
GDP and the Social Return

	(1) OLS	(2) OLS	(3) WLS	(4) OLS	(5) OLS-FD	(6) Blundell Bond (BB)	(7) BB, incl interaction S_t and S_{t-1} with time dummies
S_t	0.08320 (4.14)	0.16526 (3.13)	0.16270 (3.03)	0.02783 (0.73)	0.03995 (0.63)	0.27592 (3.24)	0.45126 (3.30)
S_t^2		-0.00747 (2.12)	-0.00736 (2.09)	-0.00074 (0.30)	0.00082 (0.22)	-0.01563 (2.85)	-0.02832 (3.70)
K_t				0.53185 (15.24)			
S_{t-1}	-0.05361 (2.71)	-0.08856 (1.63)	-0.08831 (1.59)	-0.01745 (0.43)	-0.03294 (0.62)	-0.21248 (2.52)	-0.03775 (0.40)
S_{t-1}^2		0.00396 (1.03)	0.00397 (1.03)	0.00018 (0.07)	0.00508 (1.48)	0.01525 (2.57)	0.00741 (1.39)
K_{t-1}				-0.38090 (11.31)			
y_{t-1}	0.94621 (46.29)	0.93532 (45.03)	0.93452 (44.60)	0.79246 (22.20)	0.23001 (3.46)	0.94274 (10.43)	0.64440 (4.01)
Observations	385	385	385	344	274	385	385
Nr Countries	111	111	111	95	106	111	111
F-statistic	<i>Test for joint significance of education squared and lagged education squared</i> *						
p-value		6.12	6.00	0.21	2.14	4.09	2.91
		0.00	0.00	0.81	0.12	0.02	0.01
	<i>Implied social return to education</i>						
Short run	0.08320 (4.14)	0.07120 (3.47)	0.07005 (3.39)	0.01848 (1.30)	0.05024 (1.77)	0.07916 (2.65)	0.09473 (1.14)
Long run	0.55013 (4.44)	0.50279 (4.28)	0.48486 (4.24)	0.01611 (0.60)	0.10547 (2.20)	1.02370 (0.83)	0.42262 (6.15)
	<i>Test that long run return equals short run return</i>						
F-statistic	13.46	13.85	13.78	0.01	3.79	0.57	6.35
p-value	0.00	0.00	0.00	0.92	0.05	0.45	0.01

Robust t statistics in parentheses. All columns include time dummies and a dummy for the version of the Penn World Table from where GDP data were taken. The implied return to education is calculated at an average education level of 6.3 years (the sample average in 2000). The standard errors were calculated using the delta method.

Column (7) includes interaction terms of education and education lagged with time dummies. The individual coefficient estimates are -0.03838 (0.47) in 1970, -0.02124 (0.28) in 1980 and 0.00013 (0.00) in 1990 for education, and -0.06440 (0.83) in 1970, -0.04649 (0.62) in 1980 and -0.03153 (0.46) in 1990 for lagged education. Tests for joint significance of these interaction terms are as follows. Contemporaneous interactions only: F-statistic 0.23, p-value 0.88. Lagged interactions only: F-statistic 0.26, p-value 0.86. All six interaction terms: F-statistic 2.72, p-value 0.02.

The Blundell-Bond system GMM estimator in columns (6) and (7) are 1-step estimates and assume the regressors are predetermined, not necessarily exogenous. The Hansen test of overidentifying restrictions has a p-value of 0.03 in column (6) and 0.63 in column (7). The Arellano-Bond tests for autocorrelation have the following p-values. Column (6): 0.00 for AR(1), 0.23 for AR(2). Column (7): 0.06 for AR(1), 0.30 for AR(2).

* In column (7) this test also includes the interaction terms and their lags.