Why Does Education Differ Across Countries?*

Lutz Hendricks
Iowa State University, Department of Economics
CESifo, Munich; CFS, Frankfurt
February 20, 2008

Abstract

The purpose of this paper is to investigate the reasons why average years of schooling differ across countries. A Ben-Porath model is used to decompose observed schooling differences into the contributions of demographic factors, distortions to investment in schooling and physical capital, and the skill bias of technology. The main finding is that about 80% of schooling gaps are due cross-country variation in skill bias. Distortions to school investments account for roughly 15% of the observed variation. Demographic factors and distortions to investment in physical capital make only small contributions.

JEL: I2, J24.

Key words: Education, human capital, skill bias.

1 Introduction

Educational attainment differs greatly across countries. In the data assembled by Barro and Lee (2001), average years of schooling in 1990 range from 0.9 in the five least educated countries to 11.2 in the five most educated countries. Levels accounting suggests that these educational differences account for a sizeable fraction of cross-country income gaps (Hall and Jones 1999).

The purpose of this paper is to investigate the macroeconomic forces that give rise to differences in education across countries. It complements a large literature that studies the implications of schooling differences for output per worker (e.g., Klenow and Rodriguez-Clare 1997; Hall and Jones 1999) and long-run growth (e.g., Bils and Klenow 2000; Rangazas 2000; Baier, Dwyer, and Tamura 2004). However, rather than studying the *consequences* of schooling gaps, this paper studies their *causes*. Specifically, the objective is to decompose observed cross-country differences in schooling into the contributions of education costs, demographics, and other factors suggested by theory.

Approach: These issues are studied using a version of the Ben-Porath model, which is commonly used to think about the determinants of education. Key features of the model include finitely lived households who choose education so as to maximize the present value of lifetime earnings. Competitive firms rent capital and labor services from households. Investment decisions are distorted by taxes on education and physical capital investment. In order to capture cross-country differences in skill prices, workers of different education levels are treated as imperfect substitutes in production.

The model attributes cross-country differences in education to a number of country specific parameters, which I group into four categories: demographic parameters, distortions to school investment and physical capital investment, and the skill bias of technology. The objective is

*For helpful comments I thank Peter Rangazas and seminar participants at Florida State University.

to measure the relative importance of these parameter groups for understanding cross-country schooling gaps.

In order to take the model to the data, I construct a dataset of educational attainment and its determinants covering 41 countries in the year 1990. I estimate the country specific model parameters, treating each country as a closed economy in steady state. Parameters that are common to all countries, such as preferences, are chosen to replicate U.S. data. In line with Caselli and Coleman (2006), I find that the skill bias of technologies differs dramatically across countries.

Findings: Two counter-factual experiments are used to decompose the observed cross-country schooling differences into the contributions of the four parameter groups defined above. Experiment 1 answers the question: By how much would variation in schooling be reduced, if cross-country variation in one parameter were shut down? Specifically, the experiment sets all parameters to their estimated levels, except for one, which is set to the sample average. Average years of schooling implied by the model's steady state are then compared with observed years of schooling for each country. According to this measure, by far the most important factor is the skill bias of technology; it accounts for roughly 80% of the cross-country dispersion in schooling. Distortions to school investment account for 15%, the contributions of demographic parameters are below 5%, and distortions to investment in physical capital do not contribute at all.

A second experiment yields similar results. This answers the question: By how much would schooling vary, if only one group of parameters differed across countries? Specifically, the experiment sets all parameters to their sample averages, except for one which is set to the estimated level for each country. I conclude that variation in the *demand* for skilled labor accounts for the majority of the observed schooling gaps across countries. This finding is robust to the choice of the reference country, which determines the parameters shared by all countries, and to variation in the elasticity of substitution between skill types.

The intuition stems from the fact that the fraction of skilled labor inputs varies dramatically across countries. With identical technologies, this implies differences in skill premia that are far larger than those observed in the data. Implausibly large distortions to educational investment, far greater than those implied by the data, would then be required to sustain the observed cross-country schooling gaps.

The model also sheds light on the sources of differences in skill premia across countries. Based on the same set of counter-factual experiments, I find that distortions to schooling account for the bulk of the observed variation in skill premia. Hence, the model implies a dichotomy: while the *quantity* of schooling is mostly due to technological skill bias, the *price* of schooling is mostly due to distortions to investment in education.

Perhaps the main limitation of the analysis presented here is that technology parameters are treated as exogenous. Some theories suggest that the skill bias of technology may respond to the supply of skilled labor, as in models of appropriate technologies, or to the skill premium, as in models of directed technical change.[1] Distortions to schooling could then have larger effects on educational attainment.

Literature: This paper complements a large literature on the determinants of schooling. One branch of this literature explores the quantitative implications of the Ben-Porath model, which also underlies the analysis in this paper. Bils and Klenow (2000) study the relationship between schooling and productivity growth. Rangazas (2000) examines the evolution of schooling in the U.S. over time. Kaboski (2006) studies cross-country differences in Mincer returns in a static model with a very different production structure. His model also implies that skill premia are

[1] Examples include theories of appropriate technology (Basu and Weil 1998; Acemoglu and Zilibotti 2001) or directed technical change (Acemoglu 2002).

largely determined by the cost of schooling, while educational attainment is mainly determined by the skill bias of technology. Manuelli and Seshadri (2007) study the role of schooling for cross-country income gaps. In contrast to their model, I treat workers of different skills as imperfect substitutes so that the skill premium may vary across countries.

Theories of industrialization also have implications for cross-country schooling differences. These theories highlight the importance of sectoral change from agriculture to industry (Galor and Weil 2000), income inequality and the fertility transition (Dahan and Tsiddon 1998). A related literature highlights the role of inequality for the transition to human capital intensive growth (Galor and Tsiddon 1997).

A large empirical literature studies the determinants of schooling. As pointed out by Behrman et al. (1999), much of this literature is concerned with the effects of market conditions or educational policies on *local* schooling, whereas this paper is concerned with the macroeconomic determinants of the educational attainment of entire countries.[2] Behrman et al. (1999) link schooling to aggregate economic conditions.

Finally, this paper relates to studies of the skill bias of technical change (e.g., Behrman et al. 1999) and directed technical change (Acemoglu 2002, 2003).

The paper is structured as follows. Section 2 develops the model and its calibration. The findings are presented in section 3. The intuition for the findings is discussed in section 4. The final section concludes.

2 The Model

The Ben Porath model is commonly used to study the determinants of educational attainment.[3] It features finitely lived households who choose years of schooling to maximize lifetime earnings net of education costs. In addition, the economy is populated by competitive firms who rent capital and labor from households to maximize period profits.

This paper deviates from the standard Ben Porath setup by treating workers of different education levels as imperfect substitutes. This enables the model to capture variations in skill premia across countries. The world consists of several closed economies which are in steady state. All economies are identical, except for the values of certain parameters.

2.1 Households

At each date t^* a cohort of identical agents of measure $(1+n)^{t^*}$ is born. Agents live for T periods. They start schooling at age T_E and remain in school for T_s periods, where $s \in \{1, ..., S\}$ indexes the level of education. After completing school, households work until age T_R, supplying one unit of time in each period to firms. Agents consume between the ages T_c and T.

Households maximize the discounted sum of period utilities by choosing an education level s and the age profile of consumption $c_{t,\nu,s}$:

$$\max \sum_{\nu=T_c}^{T} \beta^{\nu} \; u\left(c_{t,\nu,s}\right) \tag{1}$$

subject to budget constraint

$$a_{t+1,\nu+1,s} = R_t \; a_{t,\nu,s} + \theta_{\nu-T_E-T_s} \; w_{t,s} - m_{t,\nu,s} - c_{t,\nu,s} \tag{2}$$

[2] Rather than attempting to summarize this vast literature, I refer the reader to Behrman et al. (1999) and the references cited therein.

[3] Recent examples include Bils and Klenow (2000) and Rangazas (2000).

where ν is age and $t = t^* + \nu - 1$ denotes the date. The household owns assets $a_{t,v,s}$ which earn a gross return of R_t. Labor income is the product of a fixed experience profile of labor efficiency (θ) and the after-tax wage rate (w). While in school, the household receives no labor income ($\theta = 0$) and pays a direct education cost of $m_{t,\nu,s}$. While completing school level $\tilde{s}$, the household pays $m_{t,v,s} = \eta_{\tilde{s}}$. After completing schooling, $m_{t,v,s} = 0$. The household is not allowed to die in debt.

The optimal consumption profile satisfies the Euler equation

$$u'(c_{t,v,s}) = \beta \ R_{t+1} \ u'(c_{t+1,v+1,s}). \tag{3}$$

The optimal education level maximizes the present value of lifetime earnings net of schooling costs:

$$s = \arg\max \sum_{\nu=T_E+T_s}^{T_R} \frac{\theta_{\nu-T_E-T_s} \ w_{t,s}}{D_{t,t^*}} - \sum_{\nu=T_E}^{T_s} \frac{m_{t,\nu,s}}{D_{t,t^*}} \tag{4}$$

where D_{t,t^*} discounts date t payoffs to date t^*. Since all households are ex ante identical, all education levels must, in equilibrium, pay the same lifetime earnings net of schooling costs.

2.2 Firms

Firms rent capital, K_t, and labor of different education levels, $\mathbf{H}_t = (H_{t,1}, ..., H_{t,S})$, from households to maximize period profits.

$$\max F(K_t, \mathbf{H}_t; A_t) - q_t K_t - \sum_{s=1}^{S} w_{t,s}^f H_{t,s}$$

The production function exhibits constant returns to scale in $(K, \mathbf{H})$. A_t is a productivity factor that grows exogenously at rate γ. Profit maximizing firm behavior implies that inputs are paid their marginal products:

$$q_t = \frac{\partial F(K_t, \mathbf{H}_t; A_t)}{\partial K_t}$$
$$w_{t,s}^f = \frac{\partial F(K_t, \mathbf{H}_t; A_t)}{\partial H_{t,s}}$$

2.3 Investment Distortions

Investments in physical and human capital are distorted by *wedges*. School choice is distorted by the wedge τ_s which is modeled as a wage tax. Thus, household wages are given by

$$w_{t,s} = (1 - \tau_s) \ w_{t,s}^f. \tag{5}$$

The education wedges capture various distortions that affect school choice, such as direct costs of schooling or lost earnings due to rent seeking. Investment in physical capital is subject to the wedge ω_k, which raises the price of capital goods. The capital stock evolves according to

$$K_{t+1} = (1 - \delta) K_t + I_t / \omega_k \tag{6}$$

where and I_t denotes aggregate investment and δ is the rate of depreciation.

2.4 Market Clearing

Denote the population size by L_t and the number of persons at date t with schooling s and age v by $N_{t,v,s}$. Then the aggregate for any life-cycle profile, such as $c_{t,\nu,s}$, is given by

$$C_t = \Lambda\left(c_{t,\nu,s}\right) = \sum_{s=1}^{S}\sum_{v=1}^{T} N_{t,v,s}\ c_{t,\nu,s} \tag{7}$$

Goods market clearing requires

$$Y_t = C_t + I_t + E_t + \Omega_t \tag{8}$$

where $E_t = \Lambda\left(m_{t,v,s}\right)$ is aggregate education spending, Ω_t is the aggregate expenditure on wedges to investment in physical capital (ω_k) and human capital (τ_s). The labor market clears if

$$H_{t,s} = \sum_{\nu=T_E+T_s}^{T_R} \theta_{t,v,s}\ N_{t,\nu,s} \tag{9}$$

Finally, capital market clearing requires

$$K_t = \Lambda\left(a_{t,\nu,s}\right)/\omega_k \tag{10}$$

2.5 Competitive Equilibrium

A Competitive Equilibrium consists of an allocation

$$\{C_t, K_t, I_t, Y_t, E_t, \Omega_t, H_{t,s}, Z_t, c_{t,\nu,s}, a_{t,\nu,s}, N_{t,v,s}\}_{t=0}^{\infty}$$

and a price system $\{w_{t,s}^f, w_{t,s}, R_t\}_{t=0}^{\infty}$ such that

- The age profiles $(c_{t,v,s}, a_{t,v,s})$ maximize lifetime utility, given s.
- Households are indifferent between schooling levels.
- Factor prices equal marginal products.
- Markets clear.
- Household prices are given by

$$w_{t,s} = (1-\tau_s)\, w_{t,s}^f \tag{11}$$
$$R_t = 1 + q_t/\omega_k - \delta \tag{12}$$

- Aggregate quantities obey (7).

In what follows I consider steady state equilibria in which all variables grow at constant rates.

Irrelevance of investment wedges. An important property of the steady state is that educational attainment is invariant against the investment wedges ω_k.

Proposition 1 *Steady state educational attainment* (H_s) *is invariant against investment wedges* (ω_k).

Proof. Two features of the model are responsible for this result.

(i) The production function has the following property. If ω_k changes by factor λ and K changes by factor $\lambda^{1/(\alpha-1)}$, then labor income for each skill type $\left(w_s^f H_s\right)$, capital income (qK), and the value of the capital stock $(\omega_k K)$ all change by the same factor: $\lambda^{\alpha/(1-\alpha)}$. Thus, (12) implies that R remains unchanged. This follows from the firm's first-order conditions, which imply that w_s^f changes by $\lambda^{\alpha/(1-\alpha)}$.

(ii) The household' choice variables scale in proportion to the wage rates. That is, given R, multiplying all w_s by $\lambda > 0$ changes $c_{t,v,s}$ and $a_{t,v,s}$ by a factor of λ, while leaving the household indifferent between education levels. The reason is that the present value of lifetime earnings for each s changes by λ (recall that η_s is a multiple of teacher wages). Consumption at each age is proportional to lifetime resources by the permanent income hypothesis. The budget constraint then implies that assets also scale by λ.

Now consider an equilibrium with investment distortion $\hat{\omega}_k$ and interest rate $\hat{R}$. Imagine that $\hat{\omega}_k$ changes to $\tilde{\omega}_k = \lambda\hat{\omega}_k$. I claim that $\hat{R}$ remains unchanged in equilibrium. To show this, I construct the new equilibrium allocation. I assert that $\tilde{w}_s^f = \lambda^{\alpha/(1-\alpha)}\hat{w}_s^f$, $\tilde{q} = \lambda\hat{q}$, $\tilde{H}_s = \hat{H}_s$ and $\tilde{K} = \lambda^{1/(\alpha-1)}\hat{K}$ are the new equilibrium prices and quantities.

Due to (i), factor inputs and prices satisfy the firm's first-order conditions. Since relative wages and the interest rate are unchanged, households remain indifferent between schooling levels. Since the value of the capital stock changes by the same factor as labor income, the capital market clears. Finally, the goods market clears as all aggregate demand components as well as aggregate supply of goods scale by $\lambda^{\alpha/(1-\alpha)}$. ■

The proof reveals that the invariance of schooling against investment distortions is not a robust property of the model. It depends on the Cobb-Douglas form of the production function. In general, investment distortions have two offsetting effects on schooling. (i) The higher price of capital reduces the interest rate. (ii) Households accumulate less capital, which raises the interest rate. In this model, the two effects exactly cancel each other.

2.6 Discussion

A number of the model assumptions deserve comment.

No heterogeneity. The model abstracts from heterogeneity in the cost of schooling. This implies that the (partial equilibrium) supply of skilled labor is perfectly elastic. That is, at given prices (w_s and r), all households are indifferent between all schooling levels. A high elasticity of skilled labor supply appears consistent with empirical estimates of the effect of tuition on school attendance.

A number of studies examine how college enrollment responds to changes in tuition. Dynarski (2003, p. 286) summarizes these studies and her own findings as follows: "with the exception of the Pell studies, estimates that do and do not account for unobserved differences across individuals yield similar conclusions: a \$1,000 drop in schooling costs increases college attendance by 3 to 4 percentage points."

The following is a rough translation of Dynarski's estimates into an elasticity of college educated labor supply. Cheeseman Day & Newburger (2002) estimate that a college graduate's lifetime earnings amount to \$2.1 million (1999 dollars). These estimates are based on college graduates in the labor market between 1997 and 1999, which roughly matches the cohorts studied by Dynarski (2003). Thus, a tuition subsidy of $4,000 (reported by Dynarski in year 2000 prices) for attending four years of college increases lifetime earnings by around 0.2% and induces an increase in college attendance of around 3.5 percentage points. If half of the enrolled complete college, the implied elasticity of college labor supply with respect to the wage is around 9. Earlier studies, summarized in Dynarski (2003), imply similar elasticities. Since education is persistent across generations, the long-run effects of tuition subsidies may be even larger. Based on this calculation, I conjecture that incorporating heterogeneity in the cost of schooling into the model would not change the paper's main conclusions.

Parental investments in children. In some versions of the Ben-Porath model altruistic parents choose their children's education (e.g., Becker and Tomes 1974). If households are not borrowing constrained, parents choose the education level that maximizes their children's lifetime earnings and the model studied here emerges as a special case. Whether borrowing constraints are empirically important remains controversial (see Cameron and Taber 2004).

While my model does not explicitly permit borrowing constraints, the parameter τ_s captures any distortions that lead households to choose education levels that do not maximize lifetime earnings. In what follows, τ_s is chosen for each country such that all education levels pay the same lifetime earnings net of taxes and distortions. Thus, if borrowing constraints lead households in some countries to choose inefficiently low schooling levels, this will be captured in a positive value of τ_s.

In some versions of the Ben-Porath model parents choose fertility jointly with education (e.g., Becker and Tomes 1974). I abstract from fertility choice and treat demographic parameters as exogenous to the model. The subsequent analysis reveals that demographic parameters are not an important determinant of schooling in the model. Some evidence supporting this result is reported in section 4.1.

Structural transformation. The development literature emphasizes the structural transformation from agriculture and primary industries towards manufacturing and services (Hansen and Prescott 2002; Lucas 2004; Gollin et al. 2007). This literature suggests to think of the skill bias of technology as determined by a country's industry composition. However, Hendricks (2007) shows that most of the variation in skill bias across countries and over time occurs *within* industries. My model therefore abstracts from multiple sectors.

Endogenous skill bias. A potentially important limitation of the model is that technology parameters are exogenous. In some theories the skill bias of technology responds to the supply of skilled labor or to the skill premium. Examples are models of directed technical change (Acemoglu 2002, 2003) or multi-sector models in which industry specialization responds to factor endowments.

2.7 Model Parameters and Data

To parameterize and evaluate model, I construct a database that covers 69 countries over the period 1960 to 2000. The model is calibrated to 1990 data. This year is chosen because it covers the largest number of countries (41). The main data sources are the Penn World Table 6.2 (PWT) for national income data, Barro and Lee (2001, BL) for schooling data, and the World Development Indicators 2004 (WDI). Details on sources and variable construction are given in the Appendix. Next, I discuss how the model parameters are determined from the data.

2.7.1 Demographics

All countries share the ages at which education begins, $T_E = 6$, and at which consumption begins, $T_c = 18$. The remaining demographic parameters differ across countries. The age of retirement

T_R is taken from Kaboski (2006). The lifespan T is set to life expectancy after age 1 (WDI). The population growth rate n is set to the measured average over the period 1960 to 2000 (WDI).

2.7.2 Preferences

Households in all countries share the same preferences. For a balanced growth path to exist, households must have iso-elastic utility: $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$. The curvature parameter σ is set to a conventional value of 2. The discount factor β is chosen such that the capital output ratio for the U.S. matches the empirical value (see the Appendix).

2.7.3 Technology

Based on Gollin's (2001) findings the production function is of the Cobb Douglas form in aggregate capital K and a labor aggregator Q:

$$F(K, \mathbf{H}; A) = K^{\alpha}\ A\ Q(\mathbf{H})^{1-\alpha} \tag{13}$$

The fact that the income shares of labor vary over time and across countries requires a substitution elasticity between skill types different from unity. Following Caselli and Coleman (2006), the labor aggregator is of the Constant Elasticity of Substitution (CES) form

$$Q(\mathbf{H}) = \left\{ \sum_{s=1}^{S} \mu_s\ H_s^{\phi} \right\}^{1/\phi} \tag{14}$$

where the skill weights (μ_s) sum to unity. The value of A may be normalized to unity. Its growth rate γ is set to U.S. postwar average productivity growth, computed from the Solow residual implied by (13).

Empirical estimates place the elasticity of substitution between skilled and unskilled labor between 1.2 and 2 (see Ciccone and Peri 2005 and the references cited therein). Based on these estimates I set $\phi = 0.375$ which implies a substitution elasticity of 1.6. Lacking more detailed data, I impose the same substitution elasticity for all skill types.

The first-order condition of the firm implies

$$\frac{w_S^f}{w_1^f} = \frac{\mu_s}{\mu_1} \left(\frac{H_s}{H_1} \right)^{\phi-1} \tag{15}$$

The *skill weights* (μ_s) may thus be determined from data on wage rates and labor inputs. Labor inputs H_s are constructed according to the definition of aggregate labor supply (9). In steady state, education is independent of cohort. Hence

$$\frac{H_s}{L} = h_s \sum_{\nu=T_E+T_s}^{T_R} \theta_{\nu-T_E-T_s} \frac{N_\nu}{L} \tag{16}$$

where h_s is the fraction of the population with school level s. This is taken from BL for the population aged 15 or more. BL's data distinguish four education levels (none, primary, secondary, tertiary). This dictates the choice $S = 4$ for the model.

One problem is that the durations of school levels differ across countries. To obtain comparable definitions of education levels I proceed as follows. I set T_s to the average duration of each education level across countries. The BL data provide four points on the cumulative distribution for years of schooling. From this I estimate the fraction of persons with highest grade between T_{s-1} and T_s and use this as an estimate of h_s.

To determine skill prices (w_s^f) I draw on a database of Mincerian earnings regressions documented in detail in Hendricks (2005). A large literature estimates Mincerian equations of the form

$$w_s^f = \beta_M \ T_s + \theta_x + \varepsilon \tag{17}$$

where $x = \nu - T_E - T_s$ is experience, T_s is the duration of education, β_M is the Mincerian return to schooling, and ε is a disturbance. I use the same equation to calculate relative skill prices. Since reliable estimates of returns to experience are not available, I use a common profile for θ_x. Following Bils and Klenow (2000), this profile is given by

$$\theta_x = 0.0512 \ x - 0.00071 \ x^2 \tag{18}$$

The depreciation rate is set such that the model matches the U.S. values for I/Y and K/Y in steady state.

2.7.4 Wedges

The direct private costs of schooling (η_s) are taken from EAG (2004; see Appendix 6.1 for details). Data are available for only 22 countries. For the remaining countries, η_s is set to the cross-country average. Even for those countries where data are available measurement error is a serious concern. This problem is addressed in several ways. (i) Measurement error in η_s is subsumed in the values of τ_s. It therefore affects only the composition of school distortions into η_s and τ_s but not their total size. (ii) The sensitivity analysis in section 3.3 shows that the findings are similar when countries with missing education costs are dropped. (iii) Even setting all η_s to zero has little effect on the findings.

The distortions to physical capital investment (ω_k) are set to replicate observed capital output ratios. Finally, the distortions to education (τ_s) are set so as to make households indifferent between school levels, given observed relative skill prices (w_s) and the model's equilibrium interest rate. τ_1 is normalized to zero.

2.7.5 Summary statistics

Table 1 reports the parameters that are common to all countries. For parameters that differ across countries, table 2 shows summary statistics. A number of observations deserve mention. The discount factor β is greater than 1. This is common in life-cycle models without uncertainty. The investment wedge (ω_k) varies about four-fold across countries, which is consistent with data on the relative price of capital (Jones 1994; Hsieh and Klenow 2007).

Schooling wedges are, on average, positive. This is consistent with empirical estimates of the rate of return to schooling that are often higher than the borrowing or lending rates available to households.[4] For ease of interpretation, the direct costs of schooling (η_s) are scaled by the pre-tax wages of workers with schooling level $\hat{s} = S - 1$ (one may think of these as teacher's wages). Schooling costs are typically a small fraction of teacher's wages.

The data targeted by the calibration algorithm are shown in table 3. Mincer returns differ across countries, ranging from 0.044 to 0.154 among the sample countries. As a result, the skill premia implied by the data (w_s/w_1) vary as well.

Skill weights. Noteworthy is the very large variation in relative labor inputs (H_s/H_1) across countries. This variation is reflected in the calibrated skill weights (μ_s) shown in table 2. The μ_s vary across countries in ways that resemble the findings of Caselli and Coleman (2006). In a model of two skill types, these authors find that richer countries have higher skill weights ($A\mu_s$) for

[4] For example, Psacharopoulos and Patrinos (2002) report average private returns to schooling across countries of 8.7% for men and 9.8% for women.

Table 1: Model parameters common to all countries.

	Preferences
$\beta = 1.002$	Matches $K/Y = 2.5$
$\sigma = 2.0$	
	Technology
$\alpha = 0.33$	Gollin (2001)
$\delta = 0.045$	Matches $K/Y = 2.5$ and $I/Y = 0.18$
$\gamma = 0.018$	U.S. postwar average tfp growth rate
$\phi = 0.375$	Ciccone and Peri (2005)
$T_1 = 0$	Barro and Lee (2001)
$T_2 = 5$	Barro and Lee (2001)
$T_3 = 9$	Barro and Lee (2001)
$T_4 = 12$	Barro and Lee (2001)

Table 2: Country specific model parameters.

Demographics	Mean	Std	Min	Max
T_R	62.6	3.3	55.0	67.0
T	73.9	3.5	67.0	79.0
n (pct)	1.6	0.9	0.1	3.0
Technology				
μ_1	0.10	0.08	0.01	0.27
μ_2	0.30	0.10	0.08	0.47
μ_3	0.35	0.07	0.21	0.50
μ_4	0.26	0.12	0.07	0.58
Wedges				
ω_k	1.55	0.73	0.73	3.28
τ_2 (pct)	19.8	8.3	5.4	37.8
τ_3 (pct)	33.7	13.8	-0.2	56.9
τ_4 (pct)	38.3	16.9	-5.0	65.9
School costs				
$\eta_2/w^f_{\hat{s}}$ (pct)	2.7	2.4	0.0	9.6
$\eta_3/w^f_{\hat{s}}$ (pct)	3.4	2.7	0.0	9.5
$\eta_4/w^f_{\hat{s}}$ (pct)	25.6	24.7	0.2	89.9

Notes: The table shows summary statistics for model parameters that are country specific. Min and max are the minimum and maximum values in the sample. Std denotes the standard deviation. η_s represents the direct cost of schooling, which is scaled by the "teacher's" wage rate ($w^f_{\hat{s}}$). Countries with imputed values of η_s are omitted from the summary statistics for η_s.

Table 3: Calibration targets.

	Mean	Std	Min	Max
K/Y	2.43	0.84	1.00	4.12
I/Y	0.18	0.08	0.06	0.39
β_M	0.091	0.029	0.044	0.154
w_2/w_1	1.60	0.21	1.22	2.23
w_3/w_1	2.44	0.60	1.58	3.85
w_4/w_1	3.27	1.12	1.77	5.79
H_2/H_1	10.5	21.8	0.3	133.7
H_3/H_1	13.2	32.5	0.2	205.6
H_4/H_1	7.8	16.4	0.0	77.7
h_1	0.13	0.14	0.00	0.56
h_2	0.39	0.17	0.07	0.68
h_3	0.31	0.13	0.10	0.62
h_4	0.17	0.15	0.02	0.61

Notes: The table shows summary statistics of the observations targeted by the calibration algorithm. β_M is the Mincer return to schooling. h_s denotes the fraction of persons with schooling level s.

educated labor, but lower skill weights for uneducated labor. Similarly, the present model implies that $A\mu_1$ covaries negatively with Y/N, while the remaining skill weights covary positively with Y/N (see figure 1).

To what extent does the model account for observed differences in average years of schooling? Except for the distortions to investment (τ_s, ω_k) all of the model's country specific parameters are set independently of the modeling of school choice (though not independently of schooling data). They are either taken directly from the data (T, T_R, n, η_s) or implied by the production function together with profit maximization (μ_s, A). In order to evaluate the model's ability to account for the observed variation in schooling across countries, I therefore compute the steady state for each country while setting the τ_s to their U.S. values. There is no need to determine the investment wedges because steady state education is invariant against them.

Figure 2 compares predicted average years of schooling with 1990 data. Using the slope of an OLS regression as a measure, the model accounts for 74% of the variation in years of schooling across countries. The relationship is quite close $\left(R^2 = 0.94\right)$. I view this as an indicator that the model proposed here, in spite of its simplicity, offers a credible positive theory of educational attainment across countries.

3 Findings

3.1 Sensitivity of Schooling

In this section, the model is used to decompose the observed cross-country schooling differences into the contributions of several factors. As a first step, I explore how sensitive average years of schooling are to variation in each of the model parameters. Given information on how much each parameter varies in the data, this provides an indication which parameters could be important.

Table 4 shows the results. For each parameter, I determine its 10^{th} and 90^{th} percentile values in the data. The model's balanced growth path is computed for both values, setting all other country specific parameters are set to their cross-country averages. The table shows the resulting changes in average years of schooling.

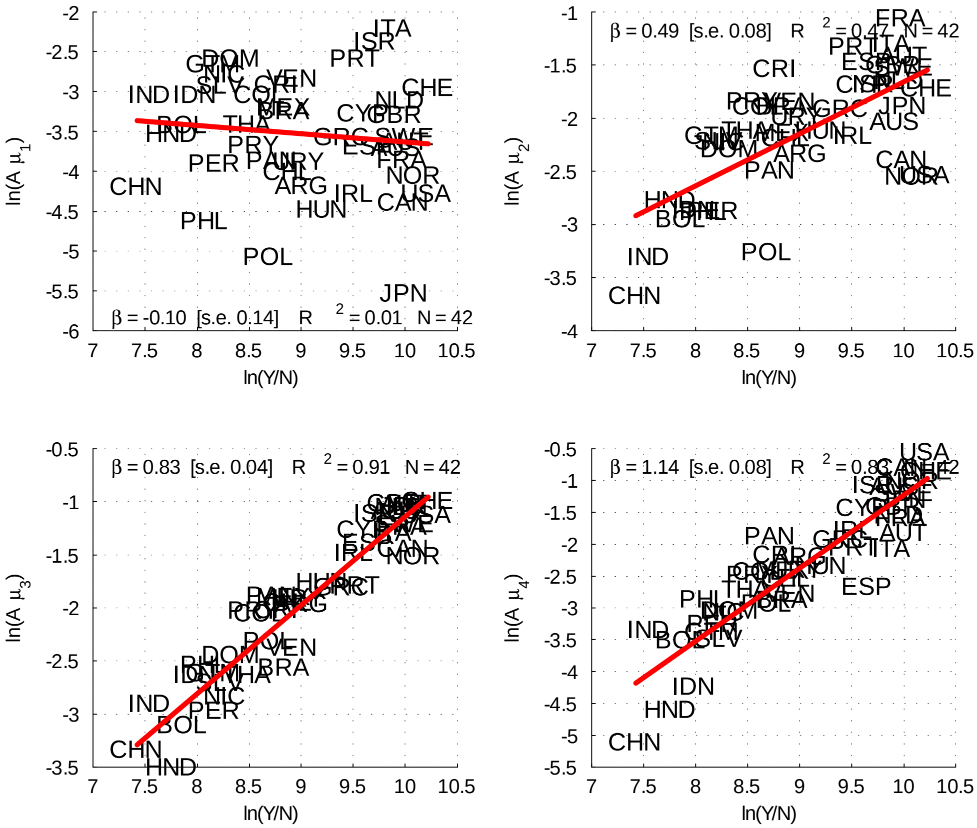

Figure 1: Skill weights and per capita GDP.

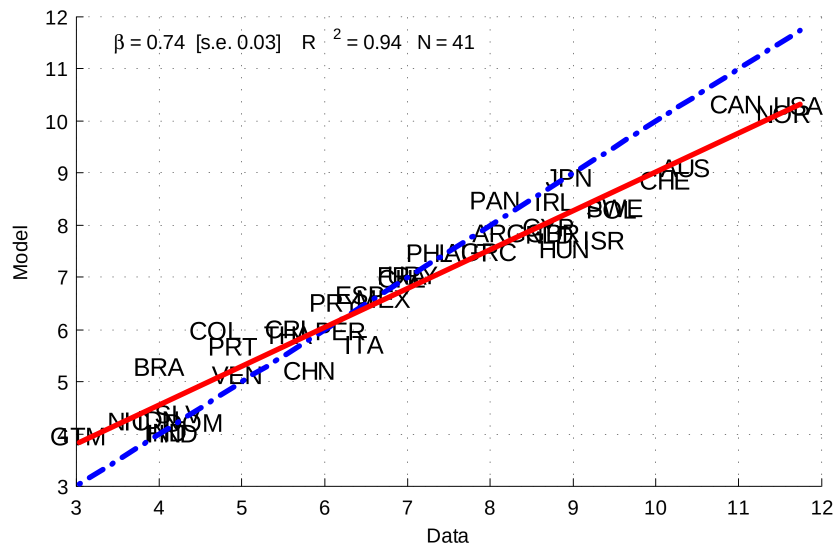

Figure 2: Years of schooling: model vs. data.

The effects of varying demographic parameters (T, T_R, n) are shown in the first four rows. Consider first the retirement age. In the data, T_R varies between 58 years (10^{th} percentile) and 65 years (90^{th} percentile). The associated steady state years of schooling are 6.8 and 6.7, respectively. Thus, years of schooling change by -0.03 for each additional year of work. Postponing retirement has two opposing effects: (i) Workers save less, which raises the interest rate and reduces the gains from schooling. (ii) The payoff period of school investments increases, which raises the value of schooling. In the calibrated model, the net effect is slightly negative.

The effects of the other demographic parameters are similarly small. Increasing the lifespan (T) raises schooling by reducing the interest rate. Higher population growth also leads to more schooling.

To assess the effects of distortions to schooling, each element of τ_s is simultaneously varied from its 10^{th} percentile value to its 90^{th} percentile value. This reduces average schooling by 1.5 years. For comparison, the cross-country gap in average schooling between the 90^{th} and the 10^{th} percentile is close to 6 years.

Varying τ_s also sheds light on the role of wage taxes. School distortions have the same effect as a wage tax whose revenues are not rebated to households: they drive a wedge between the firm's and the household's wage rate. The row labeled "τ_w" shows the effect of raising the wage tax from its 10^{th} percentile value of 22% to its 90^{th} percentile value of 48% in the dataset of Mendoza et al. (1994). Average years of schooling drop by only 0.3 years, suggesting that labor income taxation does not account for cross-country differences in schooling.[5]

Skill weights have by far the largest effect on schooling. Table 4 shows the effect of changing μ_s from Brazil's (the 10^{th} percentile country in the distribution of schooling) to Switzerland's (the 90^{th} percentile country) value. Average schooling rises by 3.2 years. This is more than half of the actual schooling gap between the two countries. Due to Proposition 1, there is no need to consider the effects of investment wedges.

[5] The effects of wage taxes on schooling could be larger in a model with endogenous work hours. Prescott (2002) and Rogerson (2006) argue that wage taxes are important for cross-country variation in hours worked.

Table 4: Sensitivity of schooling to parameter changes.

	Parameter		Schooling		Change	Slope
	10th pct	90th pct	10th pct	90th pct		
T_R	58	65	6.8	6.7	-0.2	-0.03
T	69	78	6.5	6.8	0.3	0.03
T and T_R	(58, 69)	(65, 78)	6.7	6.8	0.1	n/a
n (pct)	0.4	2.9	6.7	6.7	0.0	0.02
τ_s (pct)	(0.0, 6.4, 12.9, 14.4)	(0.0, 29.5, 52.7, 59.8)	7.3	5.8	-1.5	n/a
τ_w (pct)	22.4	47.9	7.5	7.2	-0.3	-0.01
μ_s	BRA: 4.0 years	CHE: 10.1 years	5.4	8.6	3.2	n/a

Notes: The table shows the 10^{th} and 90^{th} percentiles of each parameter in the dataset and the years of schooling implied by the model. *Slope* refers to the change in schooling divided by the change in the parameter value. The row labeled τ_w sets all τ_s to the same values, which is equivalent to a wage tax.

3.2 Accounting for Cross-country Education Gaps

The objective of this section is to quantify the contributions of various parameters to cross-country differences in educational attainment. Two counter factual experiments shed light on the issue.

The first experiment answers the question: To what extent does one parameter group by itself account for school differences across countries? The second experiment answers the question: If variation in one parameter could be eliminated, by how much would the dispersion of schooling decline? This experiment is more relevant for thinking about the effectiveness of public policies that could change, for example, the cost of schooling or distortions to investment in physical capital.

In both experiments, countries share a set of common parameters (β, σ, etc.), but differ in others, which I divide into three groups:

1. demographic parameters: T, T_R, n,

2. distortions to investment in education: η_s, τ_s,

3. technology parameters: μ_s,

Distortions to investment in physical capital (ω_k) also vary across countries but do not affect schooling.

Experiment 1. For the first experiment, all country specific parameters are set to their sample means. To measure the importance of one group of parameters, these are set to the observed or calibrated country specific values.

The results are shown in table 5. Each row represents the variation of one parameter group. To measure the contribution of each parameter, I regress the model's steady state average years of schooling on 1990 data. The slope coefficient (β) indicates how much variation in each parameter contributes to cross-country schooling gaps. The first row ("country parameters") shows the regression results when all parameters are country specific, as described in section 2.7. The model then accounts for 81% of the variation in schooling across countries.

The findings confirm the conclusions of section 3.1. Skill weights make by far the most important contribution to cross-country schooling gaps, accounting for 71% of the observed variation. Distortions to schooling are less important ($\beta = 0.12$). Demographic parameters account for less than 5% of the variation in the data.

Table 5: Experiment 1.

Experiment	β	σ_β	R^2	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Demographics	0.01	0.01	0.10	6.72
τ_s	0.12	0.03	0.25	6.64
μ_s	0.71	0.03	0.94	6.82

Notes: The table shows the results of regressing average model years of schooling on 1990 data. β denotes the regression coefficient and σ_β is its standard error. *Avg. school* shows average years of schooling across all countries implied by the model.

Experiment 2. The second experiment permits all parameters to vary across countries, except for one parameter (group) which is set to sample average values. The findings are shown in table 6. Each row again shows the results of regressing predicted against observed average years of schooling.

The results are similar to the first experiment. Eliminating skill weight differences would eliminate more than 80% of education differences. School distortions account for 10% of the observed variation. Demographic parameters account for less than 5%.

Table 6: Experiment 2.

Experiment	β	σ_β	R^2	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Avg. demographics	0.80	0.02	0.98	6.77
Avg. τ_s	0.72	0.03	0.94	6.84
Avg. μ_s	0.13	0.03	0.30	6.64

Notes: See table 5.

The main conclusion from both experiments is that skill weights account for the bulk of the cross-country schooling gaps. Figure 3 illustrates this result. It shows the model's steady state schooling when all countries share the U.S. skill weights, μ_s. All other parameters vary across countries. Eliminating variation in skill bias removes most of the observed variation in schooling. For all countries, the model predicts average schooling between 10 and 11 years, compared with 3 to 12 years in the data. In this sense, variation in skill weights is *necessary* to account for the observed cross-country differences in schooling.

3.3 Robustness

In this section, I examine the sensitivity of the findings. The *elasticity of substitution* between skilled and unskilled labor is not precisely measured in the data. Table 7 therefore repeats experiment 1 with a substitution elasticity of 4, which is at the high end of empirical estimates (see Ciccone and Peri 2005).

The findings are qualitatively unchanged. However, the importance of skill weights is reduced, while that of distortions to education is increased. Still, skill weights remain by far the most important determinant of country education.

Choosing the U.S. as the *reference country* which determines the parameters that are common to all countries is arbitrary, but has little effect on the findings. Table 8 repeats experiment 1 using Mexico as the reference country. The results closely resemble those of table 5. The same is true for a number of other reference countries.

As a final robustness check, I restrict the set of countries to those reporting data on the cost of schooling (η_s). This reduces the sample size to 22, but leaves the results of experiment 1 essentially

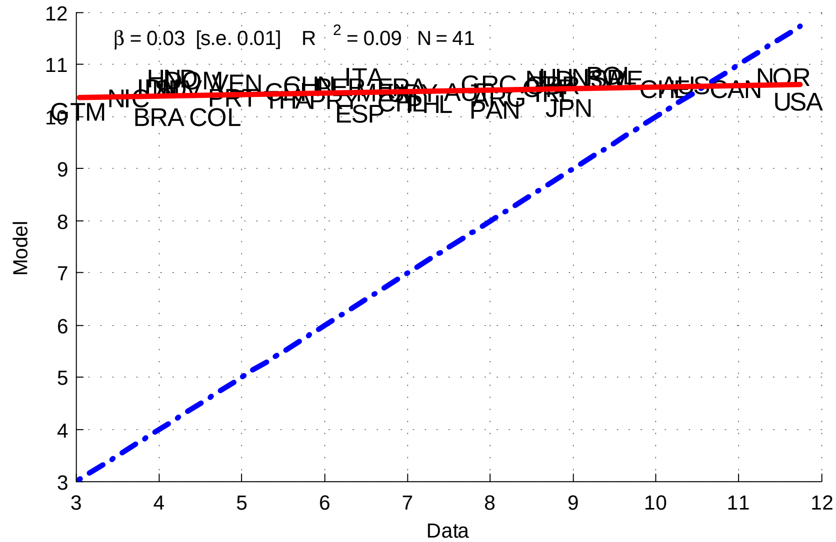

Figure 3: Average years of schooling. Common skill weights.

Table 7: Experiment 1: Substitution elasticity of skill types equals four.

Experiment	β	σ_β	R^2	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Demographics	0.05	0.01	0.34	6.94
τ_s	0.23	0.07	0.23	6.80
μ_s	0.55	0.06	0.67	6.89

Notes: See table 5.

Table 8: Experiment 1: Mexico as reference country.

Experiment	β	σ_β	R^2	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Demographics	-0.01	0.01	0.03	6.74
τ_s	0.17	0.03	0.43	6.62
μ_s	0.68	0.03	0.93	6.84

Notes: See table 5.

unchanged (see table 9).

Table 9: Experiment 1: Only countries with η_s data.

Experiment	β	σ_β	R^2	Avg. school
Country parameters	0.78	0.02	0.98	7.53
Demographics	-0.00	0.01	0.00	7.52
τ_s	0.13	0.04	0.30	7.43
μ_s	0.67	0.05	0.91	7.54

Notes: See table 5.

3.4 Why Do Returns to Schooling Differ Across Countries?

Returns to schooling differ greatly across countries (e.g., Psacharopoulos 1994). The model can be used to investigate the sources of these differences. To do so, I study the same counter factual experiments used in section 3.2 to investigate cross-country differences in the quantity of schooling.

Recall that experiment 1 allows one set of parameters to vary across countries. All other parameters are set to sample means. The results are shown in table 10, which is similar in layout to table 5 above. For each country, I estimate the Mincerian return to schooling for a sample of model households. The model's Mincer returns are then regressed against their observed counterparts. Table 10 shows the slope coefficients of this regression (β), its standard error (σ_β), and the regression R^2.

When all parameters are allowed to vary across countries (row 1), the model replicates more than 90% of the cross-country variation in Mincer returns. Almost all of this variation is due to school costs / distortions (τ_s). However, if only skill weights, demographic parameters, or investment distortions vary across countries, the model predicts that Mincer returns should be very similar in all countries. This is the mirror image of the findings for the quantity of schooling where skill weights accounted for the bulk of the observed differences.

Table 10: Experiment 1: Mincer returns.

Experiment	β	σ_β	R^2	Avg. β_M
Country parameters	0.94	0.05	0.92	0.09
Demographics	0.09	0.03	0.16	0.09
τ_s	0.91	0.05	0.90	0.09
μ_s	-0.08	0.05	0.07	0.09

Notes: The table shows the results of regressing the model's Mincer returns on their empirical counterparts. β denotes the regression coefficient and σ_β is its standard error. Avg. β_M denotes the average Mincer return across countries predicted by the model.

For experiment 2, all parameters are allowed to vary across countries. One set of parameters is then set to sample averages. This answers the question: what fraction of variation in Mincer returns would vanish, if variation in one set of parameters could be eliminated? The findings, shown in table 11, are qualitatively similar to those of the first experiment. Most of the variation in Mincer returns is again due to school costs and distortions.

Taken together with the findings reported for years of schooling, the model delivers a striking dichotomy: skill weights largely determine the quantity of schooling while distortions to school choice largely determine the price of schooling.

Table 11: Experiment 2: Mincer returns.

Experiment	β	σ_β	R^2	Avg. β_M
Country parameters	0.94	0.05	0.92	0.09
Avg. demographics	0.86	0.07	0.78	0.10
Avg. τ_s	0.01	0.04	0.00	0.09
Avg. μ_s	1.00	0.04	0.93	0.09

Notes: See table 10

4 Discussion

The intuition for the findings reported so far depends on an important feature of the data: the dispersion of labor inputs (H_j/H_i) is large relative to the dispersion of skill prices (w_j/w_i). Figure 4 illustrates this point. It shows skill premia and input ratios for several combinations of skilled and unskilled labor (i,j).

The dispersion in H_j/H_i is far too large to be consistent with a common labor demand curve across countries. For example, H_2/H_1 varies by a factor of 19.5 between the 5 most educated countries and the 5 least educated countries. By contrast, the skill premium, w_2/w_1, varies by a factor of only 1.05. Thus, a unit increase in $\ln(H_2/H_2)$ is associated, on average, with a decline in the skill premium of only around 0.02. If countries shared identical skill bias parameters, the relationship between skill premia and labor inputs would be governed by the factor demand equation

$$\ln\left(w_j^f/w_i^f\right) = \ln\left(\mu_j/\mu_i\right) - (1-\phi)\ln\left(H_j/H_i\right) \tag{19}$$

which has a slope of -0.63. A similar pattern is observed other skill pairs (i,j). Accounting for the data therefore requires large cross-country variation in demand, which is determined by the skill bias parameters $\left(\mu_j/\mu_i\right)$. Conversely, shutting down variation in skill bias would eliminate most of the variation in H_j/H_i, unless skill prices vary much more than it observed in the data. Caselli & Coleman (2006) document a similar pattern in a dataset with two skill types.

Approximate dichotomy. To understand why the model implies an approximate dichotomy where supply shifters determine skill prices and demand shifters determine skill quantities, consider an open economy version of the model. Assume that free capital mobility fixes the interest rate. Except for the market clearing condition, the definition of equilibrium is unchanged. Capital market clearing now requires that R equals the world interest rate R^*. A trade balance term is added to the goods market clearing condition.

This model permits the decomposition of cross-country differences into supply and demand factors. The demand for skilled labor is still given by the firm's first order condition (19). The remaining equilibrium conditions jointly form a "supply function." The next proposition shows that the "supply curve" is perfectly elastic.

Proposition 2 *The balanced growth wage ratios w_j/w_i for all $i,j \in \{1,...,S\}$ are invariant against changes in the skill bias parameters μ_s.*

Proof. I show that changing μ_s leads to an equilibrium in which R and q are unchanged, K,Y,Q and $C+E+\Omega$ change by a common factor λ, while w_s^f, w_s and C change by a common factor ξ. The firm's first-order condition for capital is satisfied because q is a function of K/Q. Thus, $R = 1-\delta+q/\omega_k$ is unchanged. At unchanged relative wages, there exist h_s and H_s which satisfy the firm's first-order conditions for labor inputs and labor market clearing.

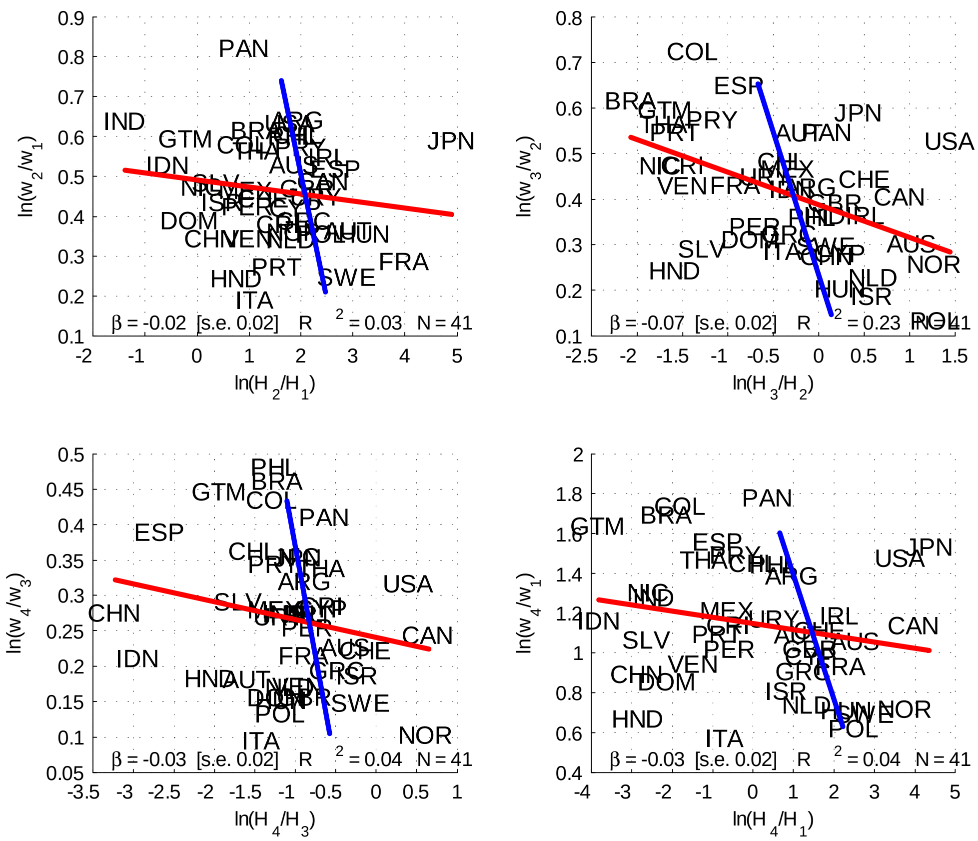

Figure 4: Skill premia and labor inputs.

If all wages change by ξ, the household remains indifferent between schooling levels. Lifetime income net of schooling costs for each s rises by ξ as does $c_{v,s}$ for all v, s. Since all households have identical age consumption profiles, C rises by ξ. Assume that households and firms borrow and lend from abroad. Aggregate asset holdings by households then change, but do not affect the supply of capital to firms.

Finally, all markets clear. Capital market clearing is implied by a constant R. Labor market clearing is automatic as households are indifferent between schooling levels. Goods market clearing is implied by Walras' law. In steady state, it requires $Y = C + \delta K + E + \Omega - RK$, where $-RK$ equals the trade balance. Thus, $Y - qK = (1 - \alpha) Y = C + E + \Omega$. The right hand side denotes household pre-tax income, which rises by λ. ■

The intuition relies on the fact that, in equilibrium, all households are indifferent between all schooling levels. Changing μ_s alters the equilibrium labor supplies H_s but not lifetime earnings. At unchanged prices, the household does not change its consumption path. As a result, all markets continue to clear. This result would no longer hold, if the interest rate were endogenous because households of different education save different amounts.

The point of Proposition 2 is that the model's balanced growth path can be represented as the intersection of a horizontal supply curve and a downward sloping demand curve for each pair of labor markets.

Figure 5 shows these supply and demand curves for fictitious countries with the average relative supplies ($\ln(H_j/H_i)$) and relative prices ($\ln(w_j/w_i)$) of the least and the most educated 10 countries. The circles represent the data points for the fictitious countries. The horizontal lines are the supply curves, which are shifted by demographic parameters and schooling costs. The downward sloping lines represent the demand curves (19). Each panel shows one pair of skill levels.

To see the interpretation of the figure, consider the first panel showing skill levels 1 and 2. The cross-country gap in H_2/H_1 may be decomposed into a demand shift component (factor 7.4) and a supply shift component (factor 1.1). The cross-country price gap w_2/w_1 may be decomposed analogously. Evidently, for all skill pairs, demand shifts account for the bulk of the quantity gaps, while supply shifts account for the bulk of the price gap.

The feature of the data which is responsible for this is the one shown in figure 4: the cross-country variation in H_j/H_i is much larger than the variation implied by moving along a fixed demand curve with an elasticity around 2.

This reasoning suggests that the elasticity of substitution between skill types is important. Figure 6 repeats the exercise with a substitution elasticity of 4. Qualitatively, this does not affect the dominant role of demand shifters for cross-country education gaps.

In this simplified model, the dichotomy is stronger than in the closed economy: skill supply is perfectly elastic and demand has no effect on prices. This raises the question how elastic the supply of skilled labor is in the closed economy. The answer evidently depends on the changes in μ_s that cause the equilibrium changes in skilled employment.

Table 12 shows supply elasticities for the following experiment. The baseline steady state assigns each country its calibrated parameters as described in section 2.7. The alternative steady state changes each country's skill bias parameters to their U.S. levels. The elasticity of "supply" is defined as $\Delta \ln(H_j/H_i) / \Delta \ln(w_j/w_i)$. The table show the cross-country average elasticity, defined as the mean of $\Delta \ln(H_j/H_i)$ divided by the mean of $\Delta \ln(w_j/w_i)$. It also shows the median elasticity.

For all skill pairs except $(i = 1, j = 2)$, the "supply" elasticities are quite large relative to the elasticities of demand. This is the reason why the intuition of the small open economy carries over to the closed economy results presented in section 3.2.

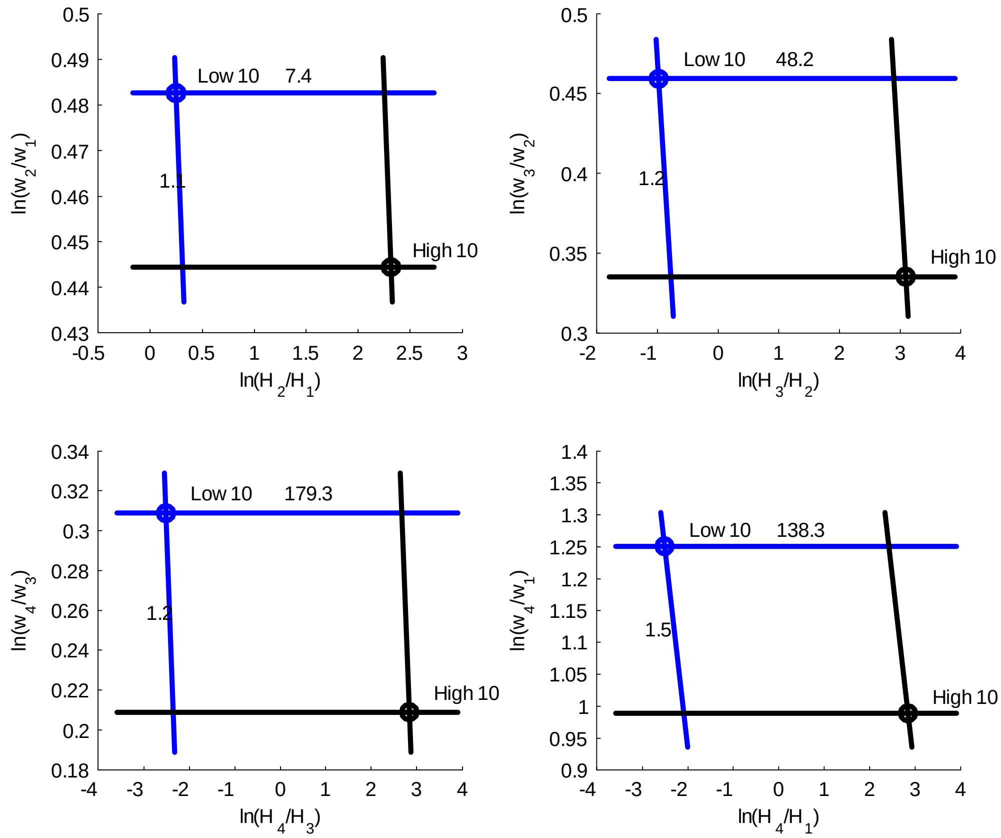

Figure 5: Labor supply and demand curves.

Table 12: Elasticities of skill "supply" functions.

Skill pair	$dln(H_j/H_i)$	$dln(w_j/w_i)$	$\frac{dln(H_j/H_i)}{dln(w_j/w_i)}$	Elasticity
(1,2)	0.39	0.128	3.0	2.3
(2,3)	1.77	0.098	18.1	20.5
(3,4)	1.41	0.083	16.9	19.4
(1,4)	3.57	0.310	11.5	12.4

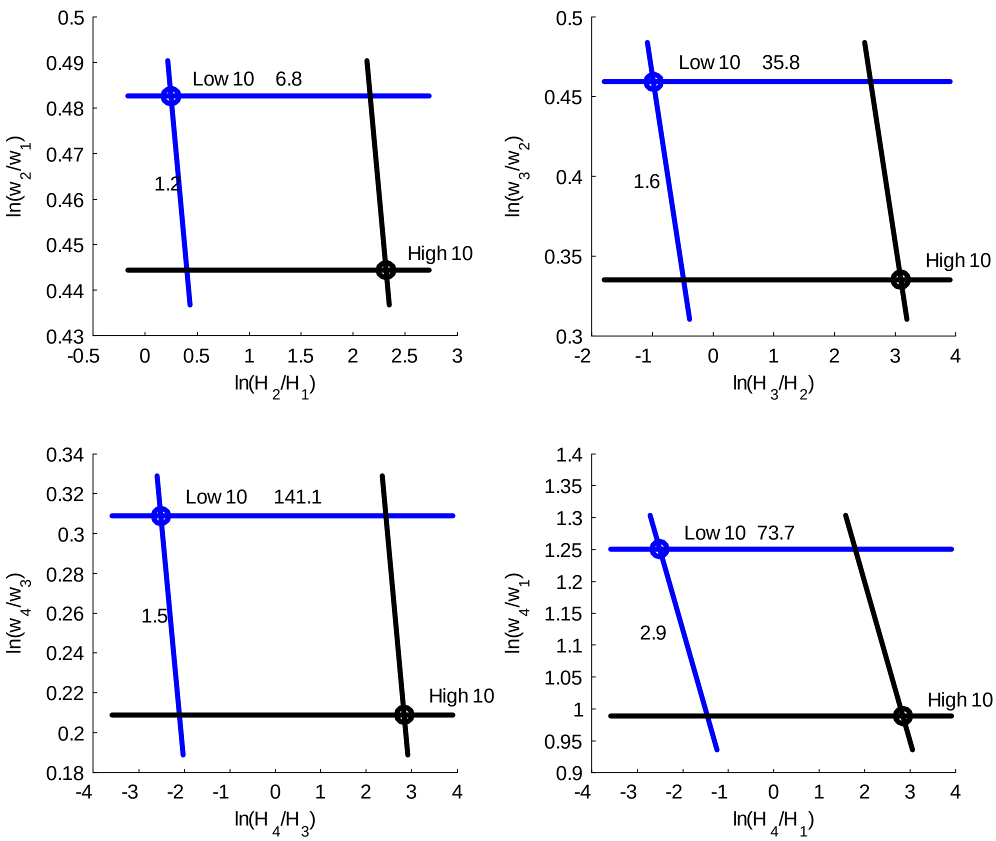

Figure 6: High elasticity of substitution.

Capital skill complementarity. One possible explanation of cross-country variation in skill bias parameters is capital skill complementarity (e.g., Krusell et al. 2000). To explore whether this is a promising explanation, consider a model with the production function

$$Y = G\left(\ell(\mathbf{L}), g\left(K, \vartheta(\mathbf{H})\right)\right)^{1-\alpha} \tag{20}$$

where $\mathbf{L} = (H_1, H_2)$ is unskilled labor and $\mathbf{H} = (H_3, H_4)$ is skilled labor. As in Krusell et al. (2000), assume a nested CES functional form:

$$G(\ell, g) = [\mu\ \ell^{\varphi} + (1-\mu)\ \ g^{\varphi}]^{1/\varphi} \tag{21}$$

$$g(K, \vartheta) = [\chi K^{\zeta} + (1-\chi)\,\vartheta^{\zeta}]^{1/\zeta} \tag{22}$$

ℓ and g are also CES aggregators. For simplicity, focus on the case where H_2/H_1 and H_4/H_3 are common across countries. It can be shown that the skill premium w_4^f/w_1^f is proportional to

$$(H_4/H_1)^{\varphi-1}\ \left[\chi\ \left(K/\vartheta\left(\mathbf{H}\right)\right)^{\zeta} + 1 - \chi\right]^{(\varphi-\zeta)/\zeta} \tag{23}$$

Capital skill complementarity means that $\zeta < \varphi$. Thus, capital skill complementarity raises the relative productivity of skilled labor in countries with high ratios of capital to skilled labor, K/ϑ. In the data, more educated countries have *lower* ratios K/H_3 and K/H_4, as shown in figure 7. Thus, capital skill complementarity does not help account for the observed cross-country variation in skill bias parameters. On the contrary; if capital skill complementarity is important, the true skill bias gaps are even larger than the calculations presented above suggest.

4.1 Demographics and Schooling

One implication of the model is that neither demographic parameters nor distortions to investment in physical capital have sizeable effects on a country's schooling level. Do the data support this model implication? To translate the model's implications into statements about correlations which can be estimated using a regression, assume that the country specific parameters are drawn from some exogenous joint distribution. The model then predicts that variations on one parameter (e.g., T) that are orthogonal to variations in μ_s and τ_s are (nearly) uncorrelated with countries' school levels.

To investigate whether this pattern of correlations holds in the data, I regress each parameter (e.g., T) on Y/N (as a proxy for μ_s) and on the Mincer return to schooling (as a proxy for τ_s).[6] I then regress years of schooling on the residual variation in T from the first regression. This approximately measures the correlation between variation in T that is orthogonal to μ_s and τ_s and years of schooling.[7]

The results are shown in table 13. To illustrate how to read this table, consider the first row referring to T_R. In the data, retirement ages vary from 59 to 65 across countries. Regressing schooling on the residual of T_R yields a coefficient of $\beta = 0.09$ with a standard error of 0.14 $\left(R^2 = 0.01\right)$. The change in schooling associated with variation in T_R is given by $(65-59)\cdot\beta$ which equals 0.52 years of schooling.

Consistent with the model's predictions, the correlations between schooling and residuals of demographic parameters are small and statistically insignificant. Varying each parameter between

[6] For the case where the dependent variable is the Mincer return, it is of course omitted from the right hand side of the regression. Alternatively, one could regress T on estimates of μ_s and τ_s in the first stage. The drawback is that these are constructed from data on educational attainment.

[7] The regression approach is valid if, as the model assumes, the parameters are exogenous. The approach is not valid if households, for example, choose fertility. However, that would be inconsistent with the model.

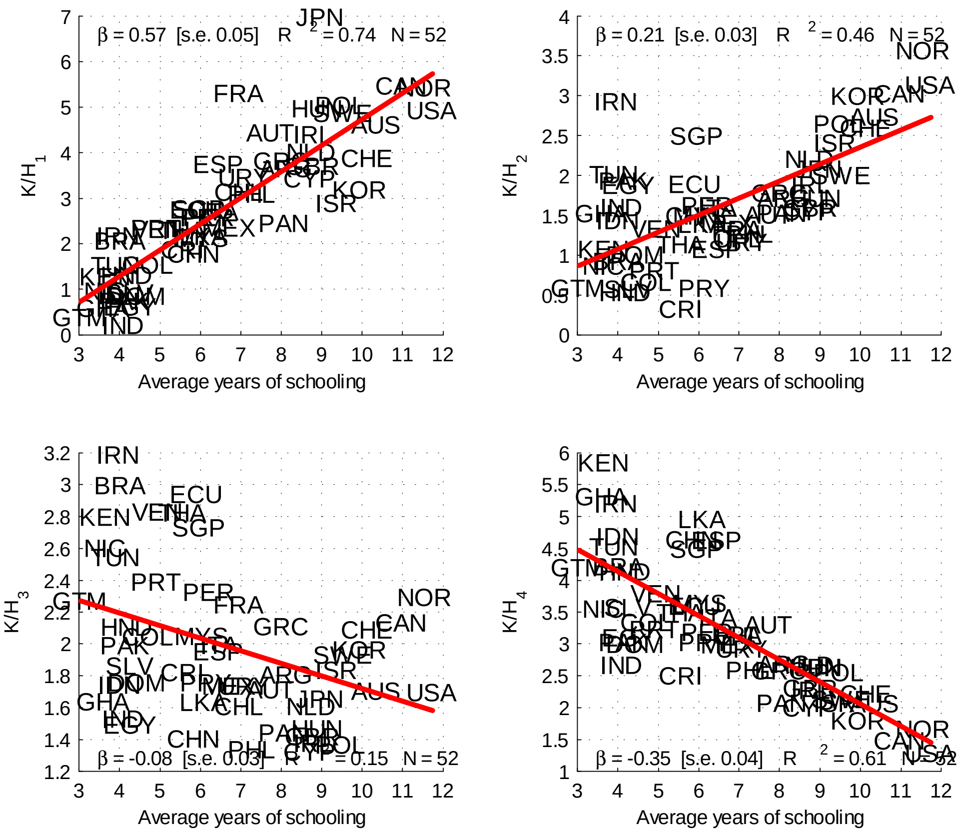

Figure 7: K/H_s and average years of schooling.

the 10^{th} percentile and the 90^{th} percentile observed in the data is associated with variations in schooling between 0.1 and 0.6 years. The only partial exception is the price of investment (proxying for ω_k) which accounts for a change in schooling of about 2 years, compared with 0 years in the model. However, this correlation is also statistically insignificant. Also consistent with the model, variations in Mincer returns (as a proxy for τ_s) account for somewhat larger school differences, but the correlation is again insignificant.

Table 13: Schooling and residual parameter variation.

	Parameter 10th pct	90th pct	Schooling change	β	σ_β	R^2
T_R	59.0	65.0	0.52	0.09	0.14	0.01
T	67.7	77.8	0.60	0.06	0.15	0.00
n (pct)	0.4	2.9	-0.07	-0.03	0.56	0.00
P_I	0.9	2.0	-2.45	-2.25	1.37	0.06
β_M (pct)	5.0	13.4	-1.53	-0.18	0.12	0.05

Notes: The table shows the results of regressing average years of schooling on the residual variation of each parameter (row). The residual variation is obtained from regressing each parameter on $\ln(Y/N)$ and the Mincer return. β is the regression coefficient and σ_β denotes its standard error. Also shown are the 10^{th} and 90^{th} percentile values of each parameter in the dataset. School change equals β times the difference between the 10^{th} and 90^{th} percentile value of each parameter.

A multiple regression of years of schooling on country specific parameters tells much the same story. Regressing schooling on Y/N alone yields an R^2 of 0.62. Adding other regressors increases R^2 to 0.71. Most of the increment is due to Mincer returns. The coefficients of demographic parameters and investment distortions are always small and statistically insignificant. These findings support the model's prediction that demographic variables have only small effects on educational attainment.

5 Conclusion

This paper studies why educational attainment differs across countries. A Ben-Porath model is used to decompose the observed cross-country differences in schooling into the contributions of demographic factors, distortions to investment in schooling and physical capital, and the skill bias of technology.

The main finding is that variation in skill bias account for around 80% of differences in schooling across countries. Distortions to school investment account for about 15%. The remaining factors, demographics and distortions to investment in physical capital, make only small contributions.

The intuition for this finding relies on the large differences in the supplies of skilled and unskilled labor observed across countries. With common technologies, these supply differences imply skill premium variations that are far larger than those observed in the data. The measured distortions to schooling are therefore too small to generate sizeable cross-country variations in educational attainment, unless more educated countries employ more skill biased technologies.

6 Appendix

6.1 Data and Variable Construction

This section describes the data sources and variable definitions. The following data sources are used:

1. Barro and Lee (2001, abbreviation BL) for educational attainment.
2. Education at a Glance (2004, EAG) for education costs.
3. Penn World Table 6.2 (PWT) for national income data.
4. World Development Indicators (2004, WDI).

6.1.1 National Income Data

National income data are taken from PWT. Table 15 summarizes the variables. The capital stock is constructed using a perpetual inventory method. The initial capital stock is set to its steady state value, $K/Y \ \ (\delta + g_Y) = I/Y$, where δ is the calibrated rate of depreciation in the model and $g_Y = 0.03$ is an estimate of the average growth rate of real GDP prior to the start of the sample.

6.1.2 Education Data

Data on educational attainment by level of schooling are taken from BL for the population aged 15 or older. Age 15 is chosen because persons of this age typically work in low income countries.

Education costs: The private costs of education (η_s) are taken from EAG (2004). Define ξ_s as the ratio of private education costs per student enrolled in schooling level s to per capita GDP. ξ_s is calculated as total expenditure per student (indicator B1) times the fraction of these expenditures borne by the private sector (indicator B3) divided by Y/N. The data are available for 2001, broken down by four education levels (pre-primary, primary, secondary, tertiary). The fraction of private spending combines primary and secondary levels. However, since these fractions are generally small, the lack of detail is unlikely to affect the findings.

In the model, I express education costs as a multiple (ψ_s) of a "teacher" wage rate, $w^f_{\hat{s}}$, so that $\eta_s = \psi_s \ w^f_{\hat{s}}$. The idea is that the education is labor intensive. One may think of households as hiring ψ_s units of teacher time while in school. To convert the observed values of ξ_s into education spending relative to teacher wages (ψ_s), I exploit the fact that the labor share is fixed $(1 - \alpha)$ and that the ratio mean earnings per person to $w^f_{\hat{s}}$ can be calculated from data on labor inputs and relative wages. Specifically, ψ_s is calculated from the identity

$$\psi_s = \xi_s \times \frac{1}{1-\alpha} \times \frac{mean\ earnings}{w^f_{\hat{s}}} \tag{24}$$

In words: Education spending in units of teacher wages equals [private education spending / per capita GDP] times [per capita GDP / mean earnings] times [mean earnings / teacher wage]. Somewhat arbitrarily I set $\hat{s} = S - 1$. This ensures that teachers are skilled but not too scarce in low income countries.

Table 14: Data by country

Country	Mincer	η_2	η_4	T_R	T	ω_k
Argentina	10.7	0.03	0.17	64	74	0.94
Australia	7.6	0.06	0.42	65	78	0.91
Austria	9.1	0.01	0.00	65	76	1.04
Brazil	15.4	n/a	n/a	65	70	1.25
Canada	8.1	n/a	n/a	65	78	0.93
Chile	12.1	0.10	0.87	65	75	1.26
China	6.9	n/a	n/a	60	72	1.74
Colombia	14.5	n/a	n/a	60	71	2.15
Costa Rica	9.5	n/a	n/a	61	77	3.26
Cyprus	7.0	n/a	n/a	65	77	1.55
Dominican Republic	7.8	n/a	n/a	60	71	2.76
El Salvador	9.7	n/a	n/a	60	70	2.94
France	7.2	0.01	0.06	60	77	1.27
Greece	6.5	0.03	0.00	65	78	0.98
Guatemala	14.9	n/a	n/a	60	67	3.28
Honduras	6.1	n/a	n/a	65	69	2.44
Hungary	5.1	0.03	0.24	60	70	1.17
India	9.1	n/a	n/a	58	67	3.18
Indonesia	10.6	0.01	0.32	55	68	2.36
Ireland	9.1	0.01	0.10	66	75	1.08
Israel	6.2	0.02	0.32	65	77	1.17
Italy	4.8	0.01	0.11	65	78	1.22
Japan	11.8	0.03	0.44	65	79	0.86
Mexico	9.3	0.02	0.19	65	74	1.35
Netherlands	5.7	0.01	0.17	65	77	1.00
Nicaragua	11.9	n/a	n/a	60	69	2.33
Norway	5.2	0.00	0.02	67	77	0.73
Panama	13.7	n/a	n/a	62	75	1.46
Paraguay	11.5	0.05	0.37	55	71	2.38
Peru	8.5	n/a	n/a	65	72	0.97
Philippines	12.0	0.07	0.45	60	70	1.54
Poland	4.4	n/a	n/a	65	72	0.88
Portugal	9.1	0.00	0.03	65	75	1.64
Spain	13.0	0.02	0.11	65	77	1.19
Sweden	5.0	0.00	0.13	65	78	1.06
Switzerland	7.4	n/a	n/a	65	78	0.77
Thailand	11.3	n/a	n/a	55	71	1.47
United Kingdom	7.8	0.04	0.20	65	76	1.28
United States	10.6	0.03	0.90	65	76	1.00
Uruguay	9.0	n/a	n/a	60	74	1.49
"Venezuela, RB"	8.6	n/a	n/a	60	73	1.27

Notes: The table shows select data by country. Mincer denotes the Mincerian return to education (β_M in [17]).

Table 15: National income variables

Symbol	Description	Variable
Y/N	Real GDP per person	Rgdpch
I/Y	Investment share	IY
C/Y	Consumption share	CI
G/Y	Government share	GI

References

[1] Acemoglu, Daron; Fabrizio Zilibotti (2001). "Productivity differences." *Quarterly Journal of Economics* 116: 563-606.

[2] Acemoglu, Daron (2002). "Directed technical change." *Review of Economic Studies* 69: 781-810.

[3] Acemoglu, Daron (2003). "Patterns of skill premia." *Review of Economic Studies* 70: 199-230.

[4] Baier, Scott; Sean Mulholland; Robert Tamura (2004). "How important are capital and total factor productivity for economic growth?" Mimeo. Clemson University.

[5] Barro, Robert; Jong-Wha Lee (2001). "International Data on Educational Attainment. Updates and Implications." *Oxford Economic Papers* 53(3): 541-63.

[6] Basu, Susanto; David N. Weil (1998). "Appropriate Technology And Growth." *The Quarterly Journal of Economics* 113(4): 1025-1054.

[7] Becker, Gary S.; Nigel Tomes (1974). "Child endowments and the quantity and quality of children." *Journal of Political Economy* 84(4), part 2: S143-62.

[8] Behrman, Jere R.; Suzanne Duryea; Miguel Szekely (1999). "Schooling investments and aggregate conditions: A household survey-based approach for Latin America and the Caribbean." Inter-American Development Bank Working Paper #407.

[9] Bils, Mark; Peter J. Klenow (2000). "Does Schooling Cause Growth?" *American Economic Review* 90(5): 1160-83.

[10] Cameron, Stephen V.; Christopher Taber (2004). "Estimation of Educational Borrowing Constraints Using Returns to Schooling." *Journal of Political Economy* 112(1) part 1: 132-82.

[11] Caselli, Francesco; Wilbur John Coleman (2006). "The world technology frontier." *American Economic Review* 96(3): 499-522.

[12] Cheeseman Day, Jennifer; Eric C. Newburger (2002). "The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings." U.S. Census Bureau Special Studies, July.

[13] Ciccone, Antonio; 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): 652-663.

[14] Dahan, Momi; Daniel Tsiddon (1998). "Demographic Transition, Income Distribution, and Economic Growth." *Journal of Economic Growth* 3(1): 29-52.

[15] Dynarski, Susan M. (2003). "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *American Economic Review* 93(1): 279-88.

[16] Galor, Oded; Daniel Tsiddon (1997). "The Distribution of Human Capital, Technological Progress, and Economic Growth." *Journal of Economic Growth* 2: 93-124.

[17] Galor, Oded; David N. Weil (2000). "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90(4): 806-28.

[18] Gollin, Douglas; Stephen L. Parente; Richard Rogerson (2007). "The food problem and the evolution of international income levels." *Journal of Monetary Economics* 54: 1230-55.

[19] Hall, Robert E.; Charles I. Jones (1999). "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, 114: 83-116.

[20] Hansen, Gary D.; Edward C. Prescott (2002). "Malthus to Solow." *American Economic Review* 92(4): 1205-1217.

[21] Hsieh, Chang-Tai; Peter Klenow (2007). "Relative prices and relative prosperity." *American Economic Review* 97(3): 562-585

[22] Hendricks, Lutz (2005). "A Database of Mincerian Earnings Regressions." www.lhendricks.org/Mincer.htm.

[23] Hendricks, Lutz (2007). "Cross-country Variation in Educational Attainment: Structural Change or Within Industry Skill Upgrading?" Mimeo. Iowa State University.

[24] Jones, Charles I. (1994). "Economic Growth and the Relative Price of Capital." *Journal of Monetary Economics* 34: 359-382.

[25] Kaboski, Joseph P. (2006). "Explaining schooling returns and output levels across countries." Mimeo. Ohio State University.

[26] Klenow, Peter; Andres Rodriguez-Clare (1997). "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" *NBER Macroeconomics Annual*. 73-103.

[27] Krusell, Per; Lee E. Ohanian; José-Víctor Ríos-Rull; Giovanni Violante (2000). "Capital-skill complementarity and inequality: a macroeconomic analysis." *Econometrica* 68(5): 1029-53.

[28] Lucas, Robert E. (2004). "Life earnings and rural-urban migration." *Journal of Political Economy* 112 (1, part 2): S29-59.

[29] Mendoza, Enrique G.; A. Razin; Linda L. Tesar (1994). "Effective tax rates in macroeconomics: cross-country estimates of tax rates on factor incomes and consumption." *Journal of Monetary Economics* 34: 297-323.

[30] Prescott, Edward C. (2002). "Prosperity and Depression." *American Economic Review* 92(2): 1-15.

[31] Psacharopoulos, George; Harry A. Patrinos (2002). "Returns to investment in education: a further update." World Bank Policy Research Working Paper 2881.

[32] Rangazas, Peter (2000). "Schooling and Economic Growth." *Journal of Monetory Economics* 46: 397-416.

[33] Rogerson, Richard (2006). "Understanding differences in hours worked." *Review of Economic Dynamics* 9: 365-409.