

# Why Does Education Differ Across Countries?

1

Lutz Hendricks, March 26, 2005

**The question:** Average schooling ranges from 0.9 to 11.2 years across countries (Barro and Lee 2000). Why?

**The purpose of the paper:** Decompose cross-country differences in schooling into the contributions of education costs, demographics, and other factors suggested by theory.

**The approach:** Study the implications of a standard Ben-Porath model.

### Key model features:

- Finitely lived households choose education so as to **maximize lifetime earnings**.
- Competitive firms rent capital and labor services from households.
- Investment decisions are **distorted** by taxes on education and physical capital investment.
- Workers of different education levels are **imperfect substitutes** in production.

The model attributes cross-country differences in education to four groups of parameters:

1. demographic parameters,
2. distortions to school investment,
3. distortions to physical capital investment,
4. the skill bias of technology (in some model versions).

**The data:** 41 countries in 1990.

Variables include:

- output per worker,
- labor and capital inputs,
- skill prices,
- direct costs of schooling,

# Summary of findings

**Result 1:** If all countries share the same [technology parameters](#), the model accounts for less than 10% of the observed variation in educational attainment.

Intuition:

- In low income countries, skilled labor is very scarce.
- With identical technologies, skill premia would be very large in low income countries.
- Distortions to education are inferred from observed skill prices and labor inputs. They are far too small to support the large skill premia implied by scarce skilled labor.

**Result 2:** With country specific skill bias parameters the [model replicates roughly three quarters](#) of the observed variation in average years of schooling.

Skill bias parameters are estimated from data on labor inputs and skill prices.

**Result 3:** The **skill bias** of technology accounts for **80%** of schooling differences across countries.

Distortions to school investment account for 15%.

Other parameters (demographics, distortions to capital investment) are unimportant.

To show this consider two **experiments**:

1. Set one parameter to sample average. All other parameters are country specific.
2. One parameter is country specific. All other parameters are set to sample averages.

For each case, compare average schooling implied by the model with the data (country by country).

**Result 4:** Most of the variation in **skill premia** across countries is due to **distortions to school investment**.

**A limitation:** Technology parameters are treated as exogenous.

In some theories, skill bias responds to the supply of skilled labor.

Examples:

- Directed technical change (Acemoglu).
- Industry specialization.
- Appropriate technology (Caselli and Coleman).

# 1 The Model

An extension of the Ben-Porath model.

Two new features:

1. Workers of different education levels are imperfect substitutes in production (to replicate skill premia).
2. The skill bias of technology may vary across countries.

The world consists of closed economies which are in steady state.

All economies are identical, except for the values of certain parameters.

## 1.1 Households

At date  $t^*$  a cohort of identical agents of measure  $(1 + n)^{t^*}$  is born.

Agents live for  $T$  periods.

Schooling starts at age  $T_E$  and lasts for  $T_s$  periods.

$s \in \{1, \dots, S\}$  indexes the level of education.

Households work until age  $T_R$ , supplying one unit of time in each period.

Agents consume between the ages  $T_C$  and  $T$ .

### 1.1.1 Household problem

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

subject to budget constraint

$$a_{t+1,\nu+1,s} = R_t a_{t,\nu,s} + \theta_{\nu-T_E-T_s} w_{t,s} (1 - \tau_s) - \eta_{t,\nu,s} - c_{t,\nu,s} + z_{t,\nu} \quad (3)$$

$\nu$  is age

$t = t^* + \nu - 1$  denotes the date.

$c_{t,\nu,s}$  is consumption at date  $t$  of a household of age  $\nu$  with schooling  $s$ .

Assets  $a_{t,\nu,s}$  earn a gross return of  $R_t$ .

Labor income is determined by

- a fixed experience profile of labor efficiency ( $\theta$ )
- the after-tax wage rate  $w_{t,s} (1 - \tau_s)$ .

While in school the household pays a direct education cost of  $\eta_{t,\nu,s}$ .

When retired, the household receives a lump-sum transfer  $z_{t,v}$ .

### 1.1.2 Optimal choices

The optimal consumption profile satisfies the Euler equation

$$u' (c_{t,v,s}) = \beta R_{t+1} u' (c_{t+1,v+1,s}). \quad (4)$$

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

$$s = \arg \max \sum_{\nu=T_E+T_s}^{T_R} \frac{\theta_{\nu-T_E-T_s} w_{t,s} (1 - \tau_s) - \eta_{t,\nu,s}}{D_{t,t^*}} \quad (5)$$

$D_{t,t^*}$  discounts date  $t$  payoffs to date  $t^*$ .

In equilibrium, all school levels must pay the same lifetime earnings net of schooling costs.

## 1.2 Firms

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

Maximize period profits:

$$\max F(K_t, \mathbf{H}_t; A_t) - q_t K_t - \sum_{s=1}^S w_{t,s}^f H_{t,s} \quad (6)$$

The production function exhibits constant returns to scale in  $(K, \mathbf{H})$ .

$A_t$  is a productivity factor that grows exogenously at rate  $\gamma$ .

Inputs are paid their marginal products:

$$q_t = \frac{\partial F(K_t, \mathbf{H}_t; A_t)}{\partial K_t} \quad (7)$$

$$w_{t,s}^f = \frac{\partial F(K_t, \mathbf{H}_t; A_t)}{\partial H_{t,s}} \quad (8)$$

## 1.3 Government

The government taxes capital and labor income at rates  $\tau_k$  and  $\tau_w$ , respectively.

Tax revenues finance government consumption ( $G_t$ ) and the lump-sum transfers to retired households ( $Z_t$ ).

The budget is balanced in every period:

$$G_t + Z_t = \tau_w \sum_{s=1}^S w_{t,s}^f H_{t,s} + \tau_r q_t K_t \quad (9)$$

The wage tax rate  $\tau_w$  differs from the household's tax rate  $\tau_s$ . The difference represents distortions to investment in schooling.

## 1.4 Market Clearing

$L_t$  is the population size.

$N_{t,v,s}$  is the measure of persons at date  $t$  with schooling  $s$  and age  $v$  by .

The aggregate for any life-cycle profile, such as  $c_{t,\nu,s}$ , is given by

$$\begin{aligned} C_t &= \Lambda(c_{t,\nu,s}) \\ &= \sum_{s=1}^S \sum_{v=1}^T N_{t,v,s} c_{t,\nu,s} \end{aligned}$$

Goods market clearing requires

$$Y_t = C_t + I_t + G_t + E_t + \Omega_t$$

where

- $E_t = \Lambda(\eta_{t,v,s})$  is aggregate education spending,
- $\Omega_t$  is the aggregate expenditure on wedges to investment in physical capital ( $\omega_k$ ) and human capital ( $\tau_s$ ),
- $I_t$  is aggregate investment.

The capital stock evolves according to

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

which defines the investment wedge  $\omega_k$ . In equilibrium,  $\omega_k$  will also be the relative price of capital.

The labor market clears if

$$H_{t,s} = \sum_{\nu=T_E+T_s}^{T_R} \theta_{t,\nu,s} N_{t,\nu,s} \quad (10)$$

Finally, capital market clearing requires

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

## 1.5 Competitive Equilibrium

A Competitive Equilibrium consists of an allocation

$$\left\{ C_t, K_t, I_t, Y_t, G_t, E_t, \Omega_t, H_{t,s}, Z_t, c_{t,\nu,s}, a_{t,\nu,s}, z_{t,\nu,s} \right\}_{t=0}^{\infty}$$

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

- The age profiles  $\left( c_{t,\nu,s}, a_{t,\nu,s} \right)$  maximize lifetime utility, given  $s$ .
- Households are indifferent between schooling levels.
- Factor prices equal marginal products.
- The government budget constraint is satisfied.
- Markets clear.
- The identities

$$w_{t,s} = (1 - \tau_s) w_{t,s}^f \quad (11)$$

$$R_t = 1 + (1 - \tau_k) (q_t/\omega_k - \delta) \quad (12)$$

are satisfied.

I consider steady state equilibria in which all endogenous variables grow at constant rates.

## 2 Model Parameters

The [database](#) covers 67 countries over the period 1960 to 2000.

Findings are reported for the year 1990 (41 countries).

The main data sources are

- Penn World Table 6.1 (PWT) for national income data,
- Barro and Lee (2000, BL) for schooling data,
- World Development Indicators (WDI).

## 2.1 Preferences

The utility function is  $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ .

$\sigma = 2$ .

$\beta$  matches  $K/Y$ .

## 2.2 Demographics

$T, T_R$  are based on 1990 data.

$n$  is post-war average population growth rate.

## 2.3 Technology

The production function is

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

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

Constant capital share  $\alpha = 0.33$  based on Gollin (2002).

Growth rate of  $A$  from post-war U.S. tfp growth (1.8% per year).

$S = 4$  is dictated by the data (no school, primary, secondary, tertiary).

Durations  $T_s$  are set to sample averages.

$\varphi$ : Substitution elasticity between  $H_s$  equal to 1.6 (Ciccone and Peri 2004).

### 2.3.1 Skill weights ( $\mu_s$ )

Observe

- relative skill prices ( $w_s^f$ )
- labor inputs ( $H_s$ ).

Labor inputs  $H_s$  are constructed according to the definition of aggregate labor supply (10).

Skill prices are estimated from Mincer equations (log wage as a function of schooling and experience).

The first-order condition of the firm implies skill weights  $\mu_s$ :

$$\frac{w_S^f}{w_1^f} = \frac{\mu_s}{\mu_1} \left( \frac{H_j}{H_i} \right)^{\varphi-1} \quad (15)$$

## 2.4 Government Policies

Tax rates are set to U.S. levels from Mendoza et al. (1994).

Transfers to retired households are 40% of mean earnings (Castaneda et al.'s 2003).

The direct costs of schooling ( $\eta_{v,s}$ ) are taken from Education at a Glance (2004).

- Interpolated for countries without data (21).
- Findings are robust against changing  $\eta$ .

Distortions to physical capital investment ( $\omega_k$ ) are set to replicate observed capital output ratios.

- $\omega_k$  varies about four-fold across countries, which is consistent with data on the relative price of capital (Jones 1994; Hsieh and Klenow 2004).

Distortions to education ( $\tau_s$ ) 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.

## 2.5 Common parameters

	Preferences
$\beta = 1.015$ $\sigma = 2.0$	Matches $K/Y = 2.7$
	Technology
$\alpha = 0.33$ $\delta = 0.036$ $\gamma = 0.018$ $\phi = 0.375$ $T_s = 0$ $T_s = 5$ $T_s = 9$ $T_s = 12$	Gollin (2001) Matches $K/Y = 2.71$ and $I/Y = 0.18$ U.S. postwar average tfp growth rate Ciccone and Peri (2004) Barro and Lee (2000) Barro and Lee (2000) Barro and Lee (2000) Barro and Lee (2000)
	Policy parameters
$\tau_w = 0.291$ $\tau_k = 0.044$	Mendoza et al. (1994) Mendoza et al. (1994)

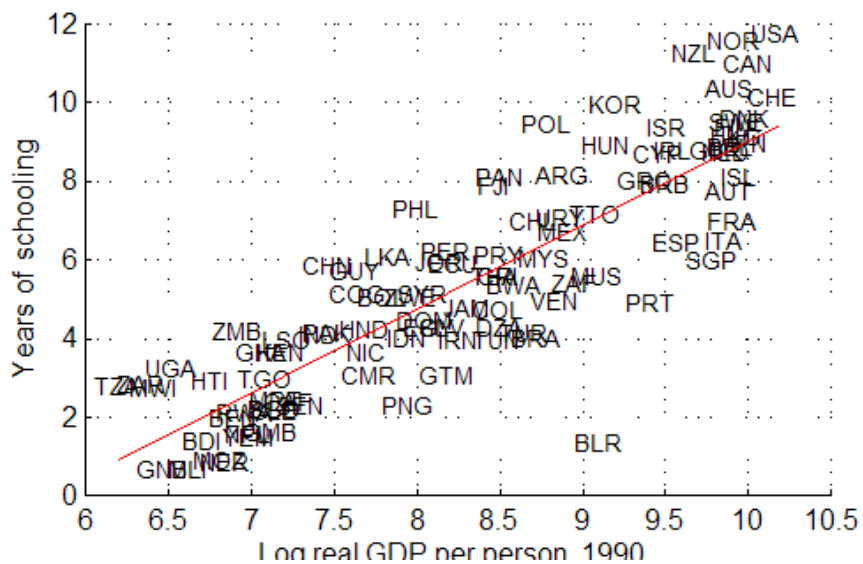
## 2.6 Country specific parameters

Demographics	Mean	Std	Min	Max
$T_R$	63.5	3.3	55.0	67.0
$T$	75.3	3.0	68.0	79.0
$n$ (pct)	1.1	0.9	0.0	2.7
Technology				
$\mu_1$	0.06	0.05	0.01	0.25
$\mu_2$	0.27	0.09	0.08	0.41
$\mu_3$	0.38	0.05	0.26	0.45
$\mu_4$	0.29	0.12	0.08	0.58
Government				
$\omega_k$	1.20	0.42	0.75	2.50
$\tau_2$ (pct)	19.6	7.5	5.8	30.2
$\tau_3$ (pct)	33.8	11.9	13.7	54.3
$\tau_4$ (pct)	38.1	14.4	13.4	64.2
$\eta_2/w_{\hat{s}}^f$ (pct)	2.0	1.8	0.0	7.2
$\eta_3/w_{\hat{s}}^f$ (pct)	2.5	2.0	0.0	7.2
$\eta_4/w_{\hat{s}}^f$ (pct)	19.2	18.6	0.1	67.2

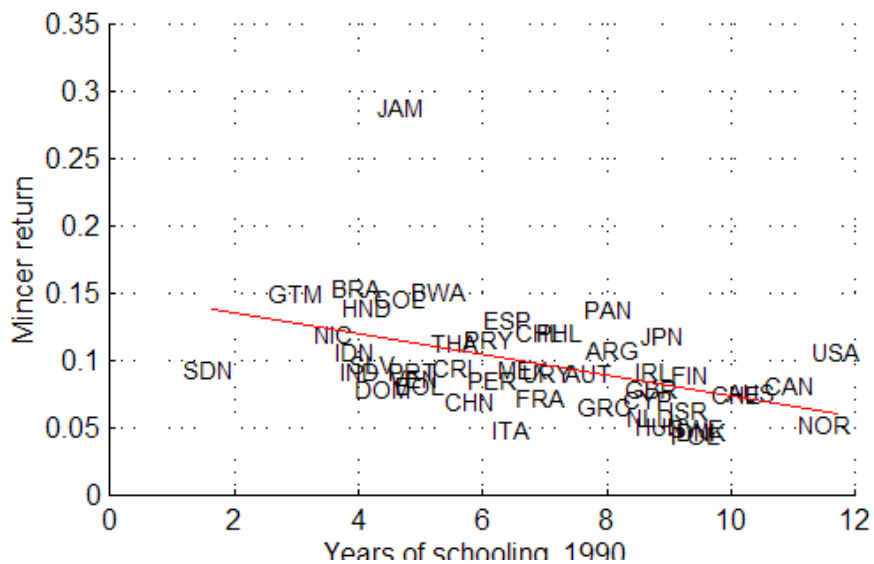
## 2.7 Calibration Targets

	Mean	Std	Min	Max
$K/Y$	2.81	0.97	1.14	4.65
$I/Y$	0.23	0.06	0.10	0.41
$\beta_M$	0.093	0.030	0.044	0.154
$w_2/w_1$	1.61	0.21	1.22	2.23
$w_3/w_1$	2.48	0.60	1.58	3.85
$w_4/w_1$	3.34	1.12	1.77	5.80
$H_2/H_1$	10.5	21.9	0.3	134.3
$H_3/H_1$	13.3	32.8	0.2	207.6
$H_4/H_1$	8.0	16.7	0.0	78.9
$h_1$	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

Note the very large variation in relative labor inputs ( $H_s/H_1$ ) across countries.



Years of schooling and real GDP.



Mincer returns and real GDP.

### 3 Findings: Common skill weights

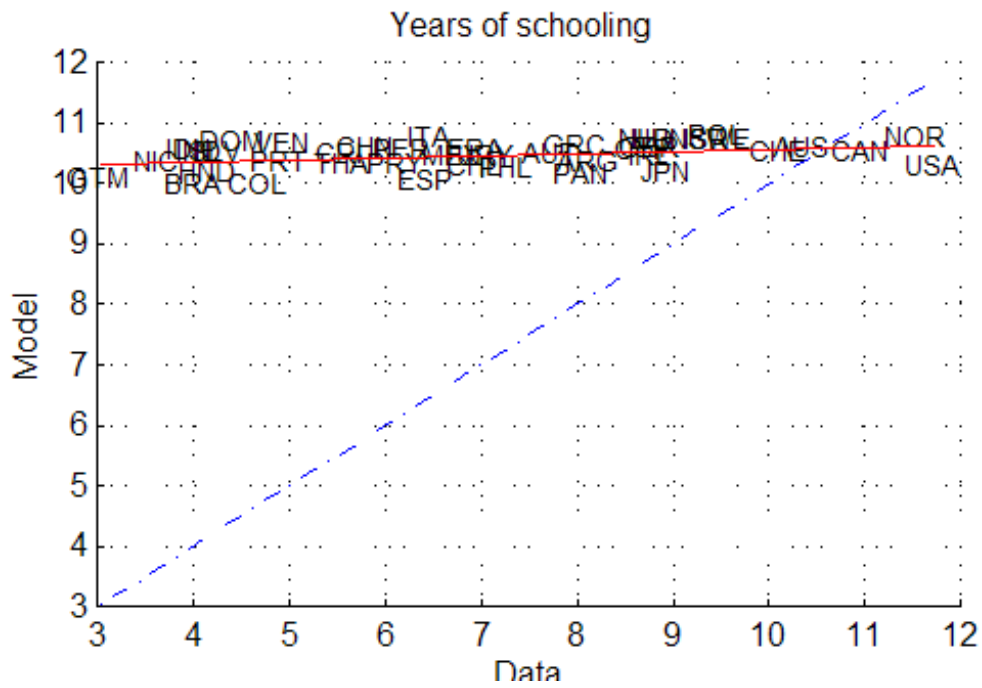
**Result 1:** With common technologies, the model accounts for less than 10% of the observed variation in schooling.

**Experiment:** Calibrate the shared model parameters, including the  $\mu_s$ , to U.S. data.

Allow other parameters to vary across countries (demographics, investment distortions, ...).

Compare the model's steady state schooling level with the data.

**Finding:** All countries have predicted schooling within one year of the U.S.



Predicted and observed years of schooling.

### 3.1 Intuition

Relative labor inputs vary dramatically across countries.

- $H_S/H_1$  for the 5 most educated countries is 31.3, compared with only 0.07 for the 5 least educated countries.
- The skill premium ranges from 1.8 to 5.8.

Skill premia are determined by marginal products:

$$\frac{w_S^f}{w_1^f} = \frac{\mu_s}{\mu_1} \left( \frac{H_j}{H_i} \right)^{\varphi-1} \quad (16)$$

With a substitution elasticity of 4,  $w_S^f/w_1^f$  should differ by a factor of 47.

The distortions implied by the data vary much less:  $\tau_s$  is at most 62%.

The model therefore poses a **trade-off**:

Either the skill premium varies too much or educational attainment varies too little across countries.

.

Note that this conclusion depends on only two features of the model:

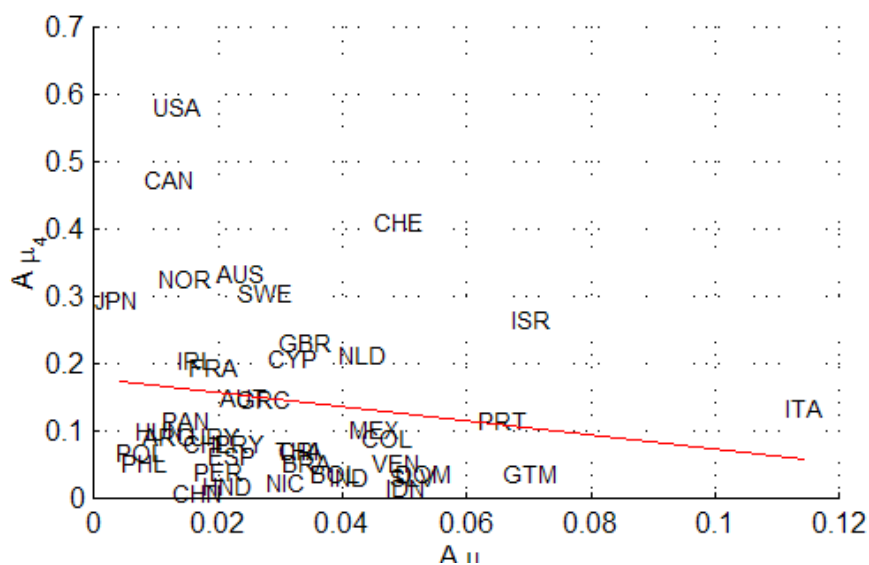
- the production function
- the assumption that workers are paid their marginal products.

## 4 Findings: country specific skill weights

From here on: Estimate  $\mu_s$  from data on  $H_s$  and  $w_s^f$ .

Patterns of skill weights resemble Caselli and Coleman (2002):

- Richer countries have higher skill weights ( $A\mu_s$ ) for educated labor, but lower skill weights for uneducated labor.
- Countries with high weights for skilled labor have low weights for unskilled labor.



Skill weights.

## 4.1 Does the model account for observed schooling?

**Experiment:** Only two model parameters are calibrated in ways that depend on the model of school choice  $(\tau_s, \omega_k)$ .

- Set those to U.S. values.

All other parameters are set as before:

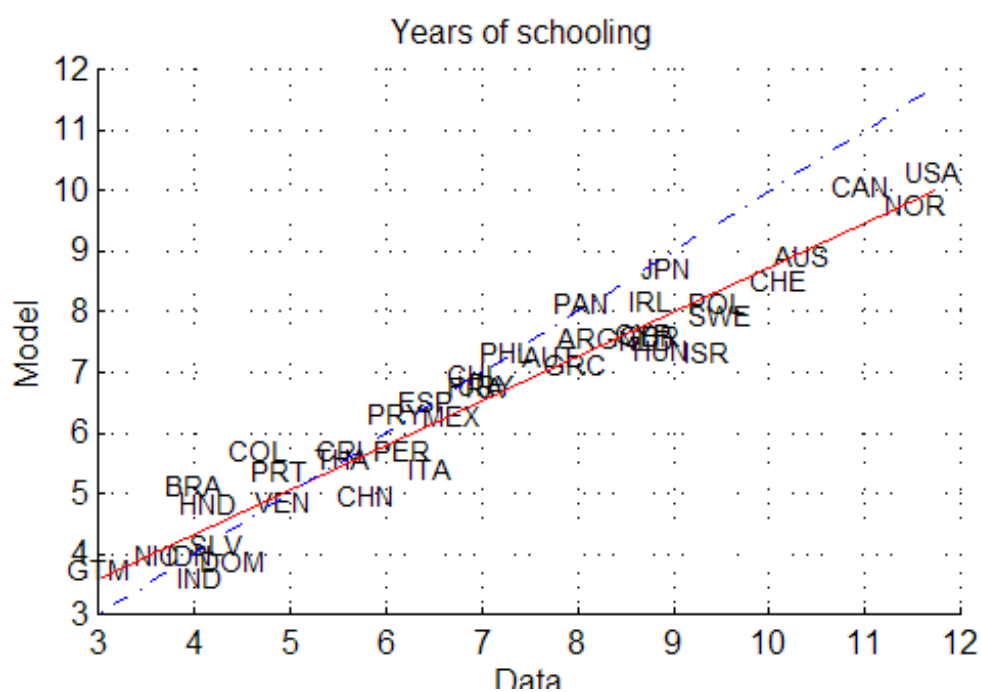
- taken directly from the data  $(T, T_R, n, \eta_s)$  or
- implied by the production function together with profit maximization  $(\mu_s, A)$ .

To **measure** the model's ability to account for observed schooling:

- Regress steady state schooling on data.
- Use the regression slope  $\beta$  as the measure.
- $\beta = 0$ : model has no predictive power.
- $\beta = 1$  and  $R^2 = 1$ : model perfectly accounts for the data.

**Result 2:** The model accounts for **73%** of the cross-country variation in schooling.

Experiment	$\beta$	$\sigma_\beta$	$R^2$	Avg. school
$\tau_s$ and $\omega_k$ common	0.73	0.03	0.93	6.61



Years of schooling: model vs. data.

## 4.2 Decomposing Education Gaps

The objective: quantify the contributions of various parameters to cross-country differences in educational attainment.

**Two experiments** answer two questions:

1. To what extent does one parameter group by itself account for school differences across countries?
2. If variation in one parameter could be eliminated, by how much would the dispersion of schooling decline?

**Experiment 1.** All country specific parameters are set to their sample means. One (group of) parameters is set to the country specific value.

Experiment	$\beta$	$\sigma_\beta$	$R^2$	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Demographics	0.01	0.00	0.09	6.70
Skill weights	0.71	0.03	0.93	6.81
S costs	0.12	0.03	0.27	6.63
Inv wedge	0.00	0.00	0.05	6.70

**Result 3:** Skill weights account for 71% of the observation variation in schooling.

School costs account for 12%.

Other parameters make small or no contributions.

**Experiment 2.** All parameters vary across countries, except for one parameter group which is set to sample average values.

Experiment	$\beta$	$\sigma_\beta$	$R^2$	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Avg. s weights	0.13	0.03	0.30	6.63
Avg. demographics	0.81	0.02	0.98	6.77
Avg. inv wedge	0.81	0.02	0.98	6.78
Avg. s cost	0.72	0.03	0.94	6.82

Findings are similar to experiment 1.

Intuition ...

**Robustness.** Repeat experiment 1 with a substitution elasticity of 4.

Experiment	$\beta$	$\sigma_{\beta}$	$R^2$	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Avg. s weights	0.28	0.06	0.32	6.79
Avg. demographics	0.77	0.02	0.98	6.78
Avg. inv wedge	0.81	0.02	0.98	6.78
Avg. s cost	0.59	0.06	0.71	6.86

**Robustness:** Repeat experiment 1 for Mexico and France as reference countries.

Experiment	$\beta$	$\sigma_{\beta}$	$R^2$	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Avg. s weights	0.15	0.03	0.37	6.63
Avg. demographics	0.82	0.02	0.98	6.77
Avg. inv wedge	0.81	0.02	0.98	6.78
Avg. s cost	0.70	0.03	0.93	6.83

Experiment	$\beta$	$\sigma_{\beta}$	$R^2$	Avg. school
Country parameters	0.81	0.02	0.98	6.78
Avg. s weights	0.12	0.03	0.28	6.63
Avg. demographics	0.81	0.02	0.98	6.77
Avg. inv wedge	0.81	0.02	0.98	6.78
Avg. s cost	0.73	0.03	0.94	6.82

## 4.2.1 Demographics and Schooling

Are model implications consistent with the data?

One implication: demographic parameters and distortions to investment in physical capital have small effects on schooling.

Varying each parameter from the 10th percentile to the 90th percentile observed in the data causes changes in average years of schooling of at most 0.2 years.

	Parameter 10th pct	90th pct	Schooling 10th pct	90th pct	Elasticity
$T_R$	58.00	65.00	6.77	6.67	-0.12
$T$	69.00	78.00	6.57	6.79	0.27
$T$ and $T_R$	58.00	65.00	6.65	6.76	0.15
$n$ (pct)	0.32	2.73	6.66	6.74	0.01
$G/Y$	0.07	0.23	6.70	6.70	0.00

The table shows the effects of varying one parameter across the range observed in the data, setting all other parameters to sample averages.

**Empirical implementation:** 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  $\mu_s$  and  $\tau_s$  are (nearly) uncorrelated with countries' school levels.

Regress each parameter (e.g.,  $T$ ) on  $Y/N$  (as a proxy for  $\mu_s$ ) and on the Mincer return to schooling (as a proxy for  $\tau_s$ ).

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  $\mu_s$  and  $\tau_s$  and years of schooling.

	Parameter 10th pct	90th pct	Schooling change	$\beta$	$\sigma_\beta$	Elasticity	$R^2$
$T_R$	59.00	65.00	0.57	0.10	0.14	0.84	0.01
$T$	67.65	77.85	0.57	0.06	0.14	0.57	0.00
$n$	0.00	0.03	0.44	17.61	59.38	0.04	0.00
$G/Y$	0.07	0.24	-0.06	-0.34	5.02	-0.01	0.00
$PI$	0.87	2.00	-2.00	-1.78	1.51	-0.32	0.03
$\beta_M$	0.05	0.14	-1.61	-18.31	12.62	-0.24	0.05

Example: retirement age  $T_R$ .

- In the data,  $T_R$  varies from 59 to 65 across countries.
- Regressing schooling on the residual of  $T_R$  yields a coefficient of  $\beta = 0.10$  (*s.e.* 0.14)  
( $R^2 = 0.01$ ).
- Variation in  $T_R$  is associated with a change in schooling of  $(65 - 59) \cdot \beta = 0.57$  years.

The findings are consistent with the model's predictions for demographic parameters and Mincer returns (as proxy for  $\tau_s$ ).

One discrepancy: In the data  $\omega_k$  is correlated with schooling (not significant).

These findings support the model's prediction that demographic variables have only small effects on educational attainment.

## 4.3 Returns to Schooling

Why do returns to schooling differ across countries?

Experiment 1 sheds light on this question:

- Allow one group of parameters to vary across countries.
- Compare Mincer returns of model and data.

Experiment	$\beta$	$\sigma_\beta$	$R^2$	Avg. $\beta$
Country parameters	0.92	0.05	0.91	0.10
Demographics	0.10	0.02	0.30	0.09
Skill weights	-0.08	0.03	0.12	0.09
S costs	0.89	0.04	0.93	0.10
Inv wedge	-0.00	0.00	0.03	0.09

Avg.  $\beta_M$  denotes the average Mincer return across countries predicted by the model.

When all parameters are allowed to vary across countries (row 1), the model replicates 90% of the cross-country variation in Mincer returns.

**Result 4:** Almost all of this variation is due to [school costs / distortions](#) ( $\eta, \tau_s$ ).

## 5 Conclusion

The model implies a **dichotomy**:

- The **quantity** of schooling is largely determined by the **skill bias** of technology.
- The **price** of schooling is largely determined by **distortions to school choice**.

**Open issues:**

- Why does skill bias vary across countries?
- Does skill bias respond to the supply of skilled labor?
- Why does educational attainment rise over time?