Explaining Schooling Returns and Output Levels Across Countries

Joseph Kaboski*

Ohio State University†

Department of Economics

March 19, 2007

Abstract

This paper develops and calibrates a model with endogenous schooling and returns to schooling to explain Mincerian returns variation across countries, and assess the importance of schooling on output variation. The calibrated model is able to explain forty percent of the variation in the data, substantially more than linear regressions explain. Variation in the direct costs of schooling driven by government funding levels relative to enrollment rates and fertility rates contribute the most to the explanatory power of the model. Nevertheless, high effective discount rates are needed to reconcile the high *level* of Mincerian returns in the data. In the calibrated model, schooling contributes on average one-third to output per worker.

*I am grateful to Virgiliu Midrigan for his exceptional work as a research assistant. I have benefited from comments from George Alessandria, Paul Evans, Jim Heckman, and Tom Mroz, as well as those received in presentations at Northwestern, Notre Dame, Ohio State, and the NEUDC 2003 and SED 2003 meetings. I thank the the Ohio State University for financial support. All mistakes are, of course, my own.

†Contact: kaboski.1@osu.edu, 415 Arps Hall, 1945 N. High St., Columbus, OH, 614-292-5588 (ph), 614-292-3906 (fax)

1 Introduction

The return to education plays an important role in the study of both development (as an indicator of schooling impacts on levels of output/worker) and inequality (as a determinant of relative wages). In development accounting, the observed returns to education are taken as given and used to measure the extent to which differences in education stocks can explain income differences across countries. In contrast, the literature on inequality has tried to model the determinants of the returns to education by focusing on secular within-country trends in these returns. This paper integrates these two approaches by: 1) developing a model with endogenous schooling and returns, 2) evaluating the models ability to explain the cross-country distribution of returns to schooling, 3) quantifying the importance of supply and demand factors in explaining this variation, and 4) evaluating the model's implications for the role of schooling in explaining cross-country income variation.

Several other papers[1] have looked toward micro-evidence on the private return to schooling to quantify schooling's role in explaining the vast disparities in income across countries. In a primary contribution to this literature, Bils and Klenow (2000) observe an inverse relationship between Mincerian[2] returns and average schooling levels in the cross-section of coun-

[1] See Bils and Klenow (2000), Hall and Jones (1999), Heckman and Klenow (1997), Klenow and Rodriguez-Claire (1997), and Krueger and Lindahl (2001).

[2] A Mincerian regression is a regression of log wages on years of schooling, controlling for experience, experience squared, and possibly other factors. The focus here is on the Mincerian return as a summary moment of the schooling/wage relationship that is available for comparison across countries. Mincer (1974) proposed an efficiency units model, in which log wages are linearly related to years of schooling, and under certain conditions the "return" (i.e. the coefficient on schooling) can be thought of as the private internal rate of return on an education investment.

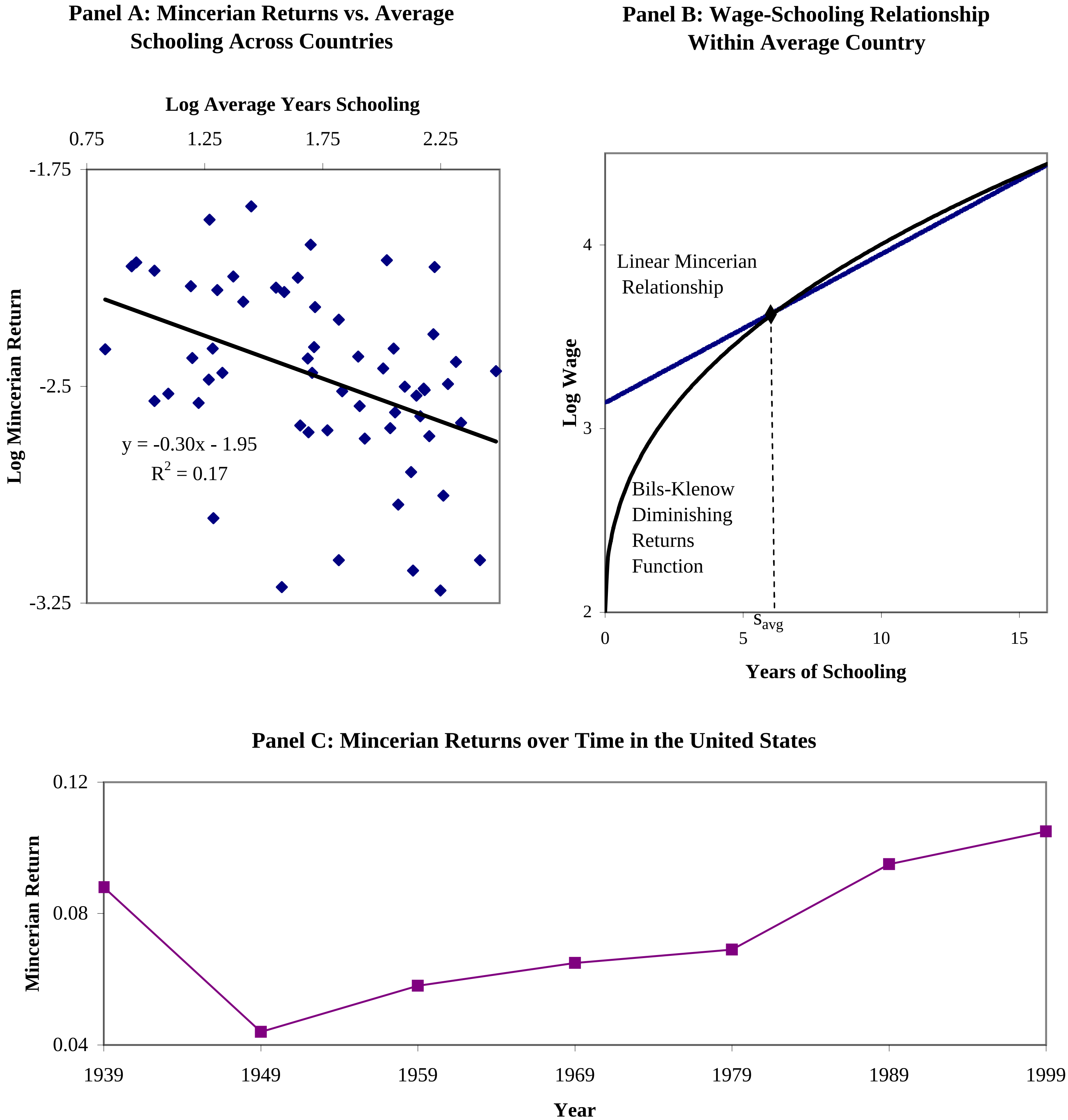
tries, and so develop a representative agent model with diminishing returns to schooling in the production of human capital efficiency units. While this approach is consistent with their cross-country observation, it is no longer consistent with two important within-country observations, namely: 1) the linear relationship in the cross-section of heterogeneous individuals within a country (i.e., the relationship from which Mincerian returns themselves are estimated), and 2) the evidence that while average schooling levels rise over time, returns to schooling fluctuate. In summary, although this literature has incorporated microevidence on the returns to schooling, the approach of modeling these returns as a technology parameterization has difficulties reconciling the within- and cross-country evidence.

The inconsistency is illustrated in Figure 1. Panel A shows the inverse relationship between Mincerian returns and average years of schooling that Bils and Klenow (BK) observed in the cross-section. The relationship is clear and significant, but explains only 17 percent of the variation in returns. Panel B shows the BK diminishing returns function relative to the linear return to schooling for the average country. Panel C shows that the returns to schooling in the United States have fluctuated over time (even over ten year periods), not trending down uniformly as schooling levels have increased.

The theory developed here is more consistent with each of the observations in Figure 1. I develop an assignment model[3] of heterogeneous workers to a range of tasks determined by the level of technology. Workers sort into these tasks based on skill-driven comparative advantage, and the labor market yields an essentially log-linear relationship between wages and schooling. The slope of this relationship, or Mincerian returns, is determined

[3] The skill-to-task assignment model builds on the work of Sattinger (1975) and Teulings (1995). This model differs by the distribution of skills.

Figure 1: Reconciling Schooling Relationships Across and Within Countries



Sources: IPUMS census data for Non-Farm Male Workers.

as an equilibrium outcome, not assumed into an efficiency units human capital production function. Consequently, the measured returns depend on factors determining the supply and demand of skill in the economy.

Though the model produces the within-country Mincerian relationship, my focus is on cross-country variation in the returns. This cross-country focus is a natural approach to study the determinants of the return to schooling given the considerable variation in the returns across countries. For example, the time series returns for the United States presented in Figure 1 ranged from 4.4 percent to 10.5 percent. In the cross-section of 59 countries[4], nearly one-third of the sample (19 of 59) have returns outside of this range.

Given the cross-country focus, the model departs in several important ways from the existing literature[5] on the supply and demand determinants of the return to schooling. First, the bulk of cross-country variation in the returns to school is long term, persistent variation. For this reason, I endogenize the supply of schooling and link this endogenous supply to the available data on factors whose variation may more plausibly be taken as structural: variation in government funding of education; fertility rates; life expectancy/expected career lengths, discount rates. Many of these factors trend over time, but again the secular movement is not quantitatively as important as the persistent cross-country variation. Furthermore, the data is not complete enough to analyze any variation over time. Hence, I abstract from dynamic considerations and apply a static model. Second, the rele-

[4] Minecerian returns are available for seventy-two countries, but only fifty-nine of these countries had all other data necessary for the analysis. The sample is slightly larger than the 52-country sample that Bils-Klenow use, but doesn't include all of their countries. I omit Cote d' Ivoire, Jamaica, Morocco, and Tanzania because of incomplete data, but include Egypt, Finland, France, Ghana, Honk Kong, Iran, Japan, Norway, S. Africa, Sri Lanka, and Tunisia.

[5] Katz and Murphy (1992), Krusell et al (2000), or Goldin and Katz (1995) are examples with exogenous schooling.

vant margin for schooling decisions varies greatly over the cross-section of countries from primary retention to tertiary completion. I therefore model schooling as a continuos choice variable in years of schooling. Finally, the assignment model of workers to tasks allows for a simple characterization of the technology-driven demand for skill in an economy. Skill is relatively more productive in frontier-technology tasks, so that a larger range of tasks translates into a higher productivity and a higher demand for skill. For each country, a skill-bias technology parameter is calibrated to match the average years of education, while a skill-neutral productivity parameter captures any remaining variation in output across countries.

After matching schooling and output, I focus on the model's predictions for Mincerian returns and the importance of education on output. The simulations are run for the full sample of countries. The key findings are:

- High effective interest rates (averaging 9.0 percent) are needed to reconcile the high observed returns to schooling, even though a large ability bias exists in the model. While many researchers have noted the puzzle of explaining high observed rates of return to schooling, this paper is the first to show it as a systematic puzzle across countries after accounting for ability bias and direct schooling costs.

- At these high interest rates, the model is able to explain about 40 percent of the variation in the data with supply factors playing the dominant role in the explained variation. Chief among these supply factors is the variation in direct costs comes from variation in school funding per student, a result of both fertility rates (16 percent of variation) and government funding rates relative to enrollment rates (19 percent). Other supply factors such as discount rates and career lengths explain very little.

- Variation in the demand for schooling, modeled as variation in the degree of skill-biased technology, plays a minor role in explaining returns to schooling, but is the most important factor in explaining levels of schooling.

- The direct contribution of schooling to income levels (i.e., the higher productivity from an educated workforce) is sizable, averaging 31 percent of income across countries. The contribution of schooling is largest in the U.S., where it is two-thirds of income levels. This is true, despite the fact that true marginal returns average only 6.0 percent (relative to 8.8 percent Mincerian returns in the data) because of a high (perhaps too high) degree of ability bias in the model.

The rest of the paper is organized as follows. The model and equilibrium equations are developed in Section 2. Section 3 describes the data and calibration methodology, and presents the estimated values of key technology parameters. The simulation results, model performance, and predictions from counterfactual simulations are discussed in Section 4, and Section 5 concludes.

2 Model

I model a competitive equilibrium in which agents choose their occupations and levels of schooling to maximize income, taking wage schedules as given. Likewise, a representative firm hires workers, taking capital and the wage schedule as given. The model is static, although the schooling decision involves a life-cycle income maximization problem. Within a country, a distribution of heterogeneous workers differ in their ability level. Across countries, economies will differ in their length of potential career (determined by entrance age, life expectancy and retirement age), levels of government expenditure per pupil (determined by levels of government funding for schooling and fertility rates), and technology levels. In order, I discuss the firm's problem, the agents' problem, equilibrium, and finally the cross-country heterogeneity in the model.

2.1 Production

A representative firm produces output using a technology that is Cobb-Douglas in capital and aggregate labor services. Capital K is treated as exogenous, but aggregate labor services L are a function of the output $x(i)$ of a continuum of imperfectly substitutable tasks indexed by their level of complexity i. These tasks are performed by a continuum of workers indexed by their skill level h[6]:

$$Y = AK^{1-\alpha}L^{\alpha} \tag{1}$$

$$L \equiv \left(\int_0^I x(i)^{1-\mu} di\right)^{\frac{1}{1-\mu}}$$

and

$$x(i) = \int_{-\infty}^{\infty} a(i,h)l(i,h)dh \text{ where } a(i,h) \equiv \tilde{a}(i,h)^{\alpha} \tag{2}$$

Here $l(i,h)$ indicates the amount of labor of human capital (or skill) level h at work in task i. $a(i,h)$ is a labor productivity parameter specific to both task and skill level, and I is the maximum complexity level of any existing task. The output $x(i)$ of a given task is the sum of the outputs of agents of different skill levels h who work in task i.

The positive function $a(i,h)$ represents the productivity that a worker of human capital level h has in task of complexity level i. It is assumed to

[6] Kaboski (2005) derives these equations from a production function with heterogeneous human capital *and* physical capital that is both task- and human capital-specific. $a(i,h)$ can thus be thought of as not only a function of human capital productivities, but also the relative productivities and prices of complementary physical capitals.

be positive and twice differentiable. In addition, I make the following three assumptions:

$$\begin{aligned}
&Assumption\ 1: \frac{\partial a(i,h)}{\partial h} \geq 0 \\
&Assumption\ 2: \frac{\partial^2 \log a(i,h)}{\partial i \partial h} > 0 \\
&Assumption\ 3: \frac{\partial^2 \log a(i,h)}{\partial h^2} < 0
\end{aligned}$$

Assumption 1 is an assumption that workers with higher levels of skills have an *absolute advantage* over less skilled workers, and ensures that wages will be increasing in skill level. Assumption 2 is a statement of *comparative advantage*; workers with higher levels of skill have a comparative advantage in performing more complex tasks. Assumption 2 is needed to ensure a sorting equilibrium, in which more skilled laborers work in more complex tasks. Assumption 3 is an assumption of *log diminishing returns*, which will allow the returns to skill to fall as the skill levels of the workforce increase.

The following parameterization will be used in the quantitative analysis:

$$a(i,h) = \exp\left(\frac{i^{1+\phi} h^{1-\phi}}{(1+\phi)(1-\phi)}\right)$$

Although the functional form appears unusual, this parameterization is differentiable, continuous, and can be easily shown to satisfy Assumptions 1-3. Furthermore, ϕ governs the importance of task relative to skill in determining productivity and also the the extent of log diminishing returns.

The firm simply maximizes profits by choosing labor taking wages $w(i,h)$ and capital K as given:

$$\max_{\substack{K \\ l(.,.)}} Y - \iint w(i,h) l(i,h) di dh$$

The firm's first-order conditions is therefore:

$$w(i,h) \geq \alpha A \left(\frac{K}{L}\right)^{1-\alpha} L^{\mu} [x(i)]^{-\mu} a(i,h) \tag{3}$$

which holds with equality for $\forall (i,h)$ such that $l(i,h) > 0$.

2.2 Households

I assume a measure-one distribution of agents differing only in a heterogeneous parameter θ, which I will refer to as ability.[7] The distribution of θ follows $g(\theta)$, a density function that is everywhere positive and finite along a bounded interval $\left[\overline{\theta}, \underline{\theta}\right]$. An agent's level of skill h is simply the sum of years of schooling s and ability θ. Taking wages and schooling costs as given, agent θ's objective is to choose a level of schooling and an occupation to maximize lifetime income net of direct schooling costs. The agents' problem is therefore:

$$\begin{aligned}
&\max_{\substack{i\in[1,I]\\ s\in[0,\bar{s}]}} \tau(s;T,r)w(i,h) - \int_0^s \eta\left[z, e(z), F\right] w(i,h)dz \\
&\text{s.t. } h = \theta + s \qquad (4)
\end{aligned}$$

Here, the expression for lifetime income $\tau(s;T,r)w(i,h)$ will be derived from an explicit lifecycle income maximization problem in the calibration section (Section 3.2). The function $\tau(.)$ represents the amount of discounted effective time spent in the labor force and will be increasing in T, the years "potential career" that can be allocated toward schooling or work, and decreasing in r, the effective discount rate of future income, and years of schooling s.

The integral in (4) represents the direct cost of schooling (i.e., tuition) born by the agent, which are proportional to indirect costs (i.e. foregone wages). The critical proportionality function $\eta(.)$ will be explicitly developed in the calibration section (Section 3.3) from a restriction that government expenditures equal government funding and students bear the remaining direct costs. The function $\eta(.)$ will vary across primary, secondary and tertiary

[7]This parameter θ could represent inherent ability, family background characteristics, or any other heterogenous quality that is a complementary to education in increasing skill or wages.

schooling levels. For given levels of schooling attainment in the economy, the direct costs function η will be decreasing in the level of government educational funding $e(s)$ at each level of schooling, but increasing in the fertility rate F which determines the relative size of the student cohort.

I assume that $w(i,h)$ is continuous and strictly concave in both i and h and later verify this in equilibrium. Given this assumption, the objective is strictly concave in i. The optimality condition for the household's choice of occupation i simplifies to:

$$\frac{\partial w(i,h)}{\partial i} = 0 \tag{5}$$

That is, given their skill level, agents work in the task that pays them the highest wage.

The first-order condition[8] for the household's choice of schooling is:

$$\frac{\partial \ln w(i,h)}{\partial h} = \frac{-\tau'(s;T,r)w(i,h) + \eta[s,e(s),F]w(i,h)}{\tau(s;T,r)w(i,h) - \int_0^s \eta\,[z,e(z),F]\,w(i,h)dz} \tag{6}$$

$$= \frac{-\tau'(s;T,r) + \eta[s,e(s),F]}{\tau(s;T,r) - \int_0^s \eta\,[z,e(z),F]\,dz} \tag{7}$$

The left-hand side shows the benefit (i.e. the increase in wages) that comes from schooling and the first right-hand side expression shows the costs. The costs in the numerator are the foregone wage $(-\tau' w)$ and the direct cost $(\eta[s,e(s),F]w)$ of a marginal increase in schooling, while the denominator is lifetime income net of schooling costs (since schooling costs also increase with the wage).

[8] In the simulations, the function $\eta(s,e(s),F)$ is a step function varying over primary, secondary and tertiary education and is therefore discontinuous over s. The function therefore contains some kinks and the first-order conditions for s do not always hold with equality (see Appendix A).

2.3 Equilibrium

Solving for an equilibrium involves applying the market clearing conditions for labor inputs (of different skill level h and in different tasks i). Assumptions 1 and 2 produce increasing mappings of tasks to abilities $\theta(i)$ and and tasks to skill levels $h(i)$.

Labor market clearing simplifies to:

$$l(i,h) = \begin{cases} \tau[s(i);T,r]g(\theta(i))\theta'(i) & \text{for } h = \theta(i) + s(i) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

In words, the demand for labor of type h working in task i must equal the supply. For task-skill combinations that satisfy $h = h(i)$, the supply is the effective time workers spend in the labor force, given their optimal level of schooling, times the density of workers of the type θ that choose task i. The $\theta'(i)$ term is the Jacobian term from transforming the density in terms of θ to a density in terms of i. For task-skill combinations that are not optimal, the supply is zero.

Given the amount of labor, the amount of task i produced is therefore[9]:

$$x(i) = a[i,h(i)]\tau[h(i) - \theta(i);T,r]g(\theta(i))\theta'(i) \tag{9}$$

Combining equation (3), the expression for wages that comes from firm optimization, with equation (5), the household optimality condition for the choice of i, yields the constant elasticity of substitution expression:

$$\frac{a_1(i,h)}{a(i,h)} = \mu\frac{x'(i)}{x(i)} \tag{10}$$

[9] Equation (2) assumed that the mass of tasks was distributed across a two-dimensional (h,i) plane. This density would need to be integrated across h in order to reduce the dimensionality to one (the i dimension). The existence of the function $h(i)$ shows that the problem was already one-dimensional, and the mass is distributed along the line $h(i)$. Hence no integration is needed.

(I use shorthand notation where $a_1(i,h)$ signifies the partial derivative of a with respect to its first element.) Taking logs and differentiating (9) and combining with (10) produces a second order differential equation in the matching function $\theta(i)$. Omitting functional dependencies, this equation is:

$$\frac{\theta''}{\theta'} + \left(\frac{g'}{g} - \frac{\tau'}{\tau}\right)\theta' + \left(\frac{a_2}{a} + \frac{\tau'}{\tau}\right)h' + \left(\frac{\mu-1}{\mu}\right)\frac{a_1}{a} = 0 \tag{11}$$

This differential equation[10] yields the optimal choice of i given θ. The corresponding optimal choice of h (and therefore s) can be easily found by applying (6). These policy functions satisfy the labor market clearing conditions by construction.

2.4 Cross-Country Heterogeneity

Though the model presented above is for a single economy, the empirical exercise will involve multiple countries. Each economy is modeled as a closed economy, i.e., using the model above, but having country-specific parameter levels (subscripted by n).

On the production side, the skill-biased technology parameter I_n represents both the amount of available tasks, and the complexity level of the most complex task. I_n is a skill-biased technology parameter and so effects both output and the demand for skill. First, *ceteris paribus,* higher values of I_n translate into higher productivity and output levels via the Romer (1990) growth effect; a larger range of tasks allow increased specialization and higher marginal products of each task.[11] Second, higher I_n values cause productivity gains from skill to be larger. This can be easily seen by looking

[10] In the quantitative simulations, η is a discontinous function of s. Hence, the mapping $i(\theta)$ and $h(\theta)$ are only piece-wise differentiable, and a series of differential equations satisfying (11) define $i(\theta)$ and $h(\theta)$.

[11] Also, the increased average productivity coming from increased average values of i for most workers is a secondary way in which higher I values translate into higher output.

at the marginal wage return to skill in the model:

$$\frac{\partial \ln w(i,h)}{\partial h} = \frac{1}{(1+\phi)} \frac{i^{1+\phi}}{h^{\phi}} \tag{12}$$

The return to skill is increasing in i. Thus, increases in I_n translate into increases in i (for all but the measure zero least skilled worker) and higher returns to skill in the economy.

The relationship between higher productivity and higher demands for skill driven by I_n is in harmony with the large empirical and theoretical literature showing that technological innovation has been skill-biased both recently[12] and across extended periods of the twentieth century.[13] Nevertheless, additional variation in technology is also needed to fully explain income levels the skill-neutral technology parameter A_n is also country-specific. It plays the role of a residual in explaining income variation not explained by the other components of the model.

Like A_n, cross-country heterogeneity in K_n affects output levels but not the decisions on schooling levels.

There are three country-specific supply-side factors in the model, all of which are explained more fully in Section 3:

1. Potential career lengths, T_n, which involves the the age of school entrance, and the minimum of either life expectancy or the retirement age;

2. The effective rate for discounting future income, r_n, which involves the real interest rate, the growth rate of wages, and the wage return

[12] See, for example, Acemoglu (2001), Berman, et al (1994), Borghans and ter Weel (2002), Caselli (1999), Greenwood and Yorukoglu (1997), Heckman, Lochner, and Taber (1998), Katz and Murphy (1992), Krueger (1993), Krusell et al (2000), and Machin and Van Reenen (1998)

[13] See, for example, Acemoglu (2000), Goldin and Katz (1998), Kaboski (2002), Lloyd-Ellis (1999), Murphy and Welch (1993), and Tinbergen (1975).

to experience. Higher interest rates imply higher discount rates, while higher growth in wages imply lower discount rates because the present opportunity cost of time is smaller relative to the future opportunity cost of time.

3. The ratio of direct to indirect costs of schooling at various levels of schooling, $\eta(s, e(s), F)$, which varies because the fraction of income spent on public educational funding $e(s)$ at different levels of schooling varies across countries and because the fertility rate varies across countries. Fertility affects direct costs because countries with high fertility rates must divide funds across a larger cohort of potential students.

The distribution of inherent ability θ, labor's share α, the diminishing return parameter ϕ, and μ, the parameter governing the elasticity of substitution between tasks, are all constant across countries.

3 Data and Calibration

This section briefly describes the data, methodology used to calibrate key technology parameters, and resulting parameter values. A more detailed description of the data and calibration methodology is given in Appendix B.

The simulations aim to match the international cross-section of economies in 1990, chosen because it was the year of best data availability. Although the model is static, because of trends in the data and the implicit timing in the model, relevant years for data must be chosen. For example, I use fertility and life expectancy data averaged over 1960-1970, since I am concerned with members of the labor force in 1990. Details of these decisions are again included in Appendix B. The target Mincerian returns are also averages over the period 1975 to 1995. The panel variation in these data is small relative to the cross-sectional variation.

Data variables in this dataset[14] play one of three roles in the analysis, as either a direct input into the model as a parameter value, a target variable for estimating technology parameters, and/or a variable for evaluating the model's predictions. Table 1 presents the summary statistics of the crucial data variables used.

The calibration involves four main components: 1) the distribution of ability $g(\theta)$, which is common to all countries; 2) effective working time $\tau(s;T,r)$ as a function of potential career, schooling, and an effective discount rate; 3) the ratio of direct to indirect costs of schooling at various levels of schooling, $\eta(s,e(s),F)$, as a function of government expenditures on education at different levels, school enrollment rates and fertility; and, finally, 4) the common and country-specific technology parameters. Table 2 summarizes the calibrated parameters, targets and values.

3.1 Distribution of Ability

The distribution of ability $g(\theta)$ is assumed constant across all countries. Given the human capital production function, inherent ability should be measured in year of schooling equivalents. Using an AFQT test score proxy for ability, Cawley, Heckman and Vytlacil (1999) estimate that in log wage terms, the gain from being in a higher ability quartile is about 1.5 times the gain from an additional year of school in the United States.[15] Since the available evidence is on ability quartiles, the uniform distribution of ability

[14] The cross-country dataset, original data sources, and data programs are available at http://kaboski.econ.ohio-state.edu/ccmincerdata.html.

[15] Their results range from 1.3 to 2.3 for different race-gender groups. Since the numbers for white males and white females – the bulk of the labor force – are 1.4 and 1.4 respectively, I calibrate to a value of 1.5, which is intermediate but close to the mode.

Table 1: Data Summary Statistics

Variable	Mean	Standard Deviation	Coeff. Of Variation	Min	Max
Avg. Years of Schooling	6.2	2.5	0.40	2.3	12
Avg. Mincerian Return	8.8%	2.6%	0.30	4.1%	15.3%
ln GDP/Worker	9.4	0.8	0.08	7.5	10.5
Capital/Output	2.1	0.8	0.37	0.5	4.5
Primary Funding (% of GDP)	1.8%	0.8%	0.43	0.4%	4.1%
Secondary Funding (% of GDP)	1.4%	0.8%	0.55	0.2%	3.3%
Tertiary Funding (% of GDP)	0.4%	0.2%	0.55	0.1%	1.0%
Total Funding (% of GDP)	3.7%	1.4%	0.39	0.7%	6.6%
Total Fertility Rate	4.3	1.6	0.36	1.9	7.2
Life Expectancy	64.6	7.3	0.11	50.1	74.0
Potential Career Length	54.2	4.8	0.09	43.1	61.0
Effective Discount Rate	5.1%	0.3%	0.07	4.2%	6.3%
ln A (Skill Neutral Tech. Parameter)	4.3	0.6	0.14	2.7	5.4
I (Skill-Biased Tech. Parameter)	1.09	0.12	0.11	0.87	1.44

Notes: Fertility and life expectancy have been adjusted for infant mortality. Fertility data are divided by two to yield F (offspring per capita). Effective discount rates are for ρ=0.066, and I and A are the calibrated values for ρ=0.066 and μ=0.7.

Table 2: Calibration Summary		
Parameter	**Value**	**Calibration Target**
Ability Distribution		
$\underline{\theta}$	0.01*	Avg. 75th-25th percentile difference in schooling levels in the cross-section of countries
$\overline{\theta}-\underline{\theta}$	6*	Return to ability relative to schooling*
Effective Working Time		
ρ	0.04/ 0.067	Standard discount rate/ Average mincerian returns in the cross-section of countries
x	0.015	Avg. linearized return to experience in available countries
σ	1	Standard intertemporal elasticity of substitution
Direct Schooling Costs		
$\tilde{\eta}_j$	0.13, 0.30, and 0.42	Average $\tilde{\eta}_{j,gov}$ of countries that fully fund schooling at level j (i.e., primary, secondary, tertiary)
Technology		
α	0.66	Labor's share
μ	0.7*	Inverse elasticity of substitition between college and high school workers
I_n	various, see Appendix E	Average schooling in country n
A_n	various, see Appendix E	Output per worker in country n
ϕ	0.76*	Minimized variation in ln A

* Results were examined for robustness to these values.

θ was used with a range of ability equal to six (i.e. $\overline{\theta} - \underline{\theta} = 4 * 1.5 = 6$). The lower bound, $\underline{\theta} = 0.01$, was chosen to match the model's prediction of the average difference in years of schooling between the 25th and 75th percentile to the cross-country average of 5.9 years in the data. This calibration yields what I consider an upper bound of plausible levels of ability bias in Mincerian returns[16], and is viewed as a conservative calibration in light of the major claims of Section 4.

3.2 Effective Working Time

Effective working time is a function of years of schooling, potential career length, and a discount rate. Potential career length T is calculated as the difference between min(life expectancy, retirement age) and primary school entrance age.

A typical lifecycle approach would express discounted lifetime earnings as:

$$\int_s^T e^{-\tilde{r}(t-s_{avg})} w(s+\theta; s) e^{\gamma(t-s)} e^{x(t-s)} dt$$

where γ captures growth in wages over time, ξ captures a linear[17] return to experience, and discounts earnings at a rate $\tilde{r}$. I use average schooling

[16] The variation in ability, and consequently the contribution of ability to Mincerian returns (i.e., the ability bias) in the model may too large for three reasons. First, many people believe that AFQT test scores are themselves affected by schooling. Second, given a positive correlation between schooling and ability, the bias in any given country will be governed by the range of schooling in the country. In the data, a cross-country regression of Mincerian return on interquartile schooling range produces a significant coefficient of -0.0040, but in the model this coefficient is -0.0076. Third, instrumental variables estimates of returns to schooling are usually higher than Mincerian returns, though this may be explained by heterogeneity across individuals (see Heckman and Vytlacil, 2004).

[17] My omission of a quadratic return to experience departs from the Mincerian model, but allows a closed form expression for τ. This omission is not viewed as a crucial problem. Given discounting, the negative returns to experience at the end of careers is not quantitatively important. I do, however, evaluate the robustness of results to variation in

level, s_{avg}, as a point of reference for discounting, since this is when the average worker is on the margin between more schooling and entering the labor market. Comparing the above expression to lifetime wage earnings in (4), would yield:

$$\tilde{\tau}(s;T,r,s_{avg}) = \frac{e^{-r(T-s_{avg})} - e^{-r(s-s_{avg})}}{-r} \tag{13}$$

where $r = \tilde{r} - \gamma - x$. The discount rate $r_n = \tilde{r}_n - \gamma_n - x$ incorporates a country-specific real interest rate, $\tilde{r}_n$ and growth in wages, γ_n, and a common return to experience x common across all countries[18].

The growth in output per worker (1960-1990) is used as a measure of γ_n. I use the real interest rate $\tilde{r}_n$ implied from the neoclassical growth model's Euler equation given the growth rate of consumption (income per equivalent per capita, 1960-1990) in the data.[19] Because the non-linearity of $\tilde{\tau}$ precludes analytical representation of $h(\theta)$ needed to solve the model,

---

$(\gamma + x - r)$.

[18] Returns to experience were not available for all countries, but the linearized return averaged 1.5 percent across all countries with available estimates.

[19] The Euler equation from a neoclassical growth model implies:

$$\gamma_{c,n} = \frac{1}{\sigma}(\tilde{r}_n - \rho)$$

I calibrate the intertemporal elasticity of substitution σ to be to be unity, and therefore:

$$\tilde{r}_n = \gamma_{c,n} + \rho$$

We assume the discount rate to be common across all countries. For a standard value of $\rho = 0.04$, the implied discount rates r_n are too small to reconcile the high level of Mincerian returns in the data. A higher discount rate ($\rho = 0.067$) is needed to produce r_n values high enough to match the average Mincerian return in the data. I view these high effective discount rates as likely the result of credit market imperfections.

τ is a linear approximation[20] of $\tilde{\tau}$ around s_{avg}.

3.3 Direct Schooling Costs

The ratio of direct to indirect schooling costs $\eta(s, e(s), F)$ paid by students is calibrated as a step function varying across primary, secondary, and tertiary education. The data dictates the length of schooling at each level, which varies across countries. The direct costs paid by students are calibrated as the difference between true direct costs relative to indirect costs $\tilde{\eta}$ (common to all countries) and $\tilde{\eta}_{j,gov}$, which is paid by the government (country-specific).

The calculation of $\tilde{\eta}_{j,gov,n}$ involves dividing total government expenditures (at level j) by total indirect costs (at level j). Similar to Kaboski (2001), I assume that the government designates a fraction e_j of old generation output and divides it among the young generation.[21] Variation in $\tilde{\eta}_{j,gov,n}$ is therefore driven by two factors: the size of the young cohort relative to the old (i.e., fertility F_j) and the level of government spending (i.e., e_j) relative to the enrollment rates at different levels of schooling.

The ratio of direct costs paid by the government to indirect costs is equal to the total government funding at level j ($e_j Y_j$) divided by the total direct

[20] Thus $\tau\ (s; T, r, s_{avg})$ is the linear approximation of $\tilde{\tau}$ around s_{avg}:

$$\begin{aligned} \tau(s; T, r, s_{avg}) &= c_1 + c_2 s \\ c_1 &\equiv \frac{e^{-r(T-s_{avg})} - 1}{-r} + s_{avg} \\ c_2 &\equiv -1 \end{aligned}$$

This linearization has the convenient property of having the foregone time cost of a year of schooling equaling one, a second justification for using s_{avg} as the point of reference for discounting.

[21] The determination of this fraction, indeed the source of this bequest, is not described in the model above. In order to allow this to be mapped into available data, I assume that the funding is a bequest from an unmodeled previous period.

costs of schooling (which depends on costs of schooling, number of students, and years in school of each student) multiplied by the ratio of direct to indirect costs. The formula is[22]:

$$\tilde{\eta}_{j,gov,n} \equiv \frac{e_{j,n} Y_{j,n}}{F_n \int s_{j,n}(\theta) w_n(\theta) g(\theta) d\theta} \tag{14}$$

where $s_{j\,n}$ is the number of years of schooling at level j.

I calibrate the true cost of schooling $\tilde{\eta}_j$ using the calculated values of $\tilde{\eta}_{j,gov\,n}$ as the average value of the subsidy level $\tilde{\eta}_{j,gov\,n}$ across countries that fully funded education at level j. For primary and secondary, fully funded countries were those with compulsory schooling at each level, while for tertiary schooling I used a subset of countries with no tertiary school tuition and at least 90 percent of total tertiary expenditures were publicly funded. The estimated true costs of schooling $\tilde{\eta}_j$ as a fraction of forgone earnings 0.13 (primary), 0.30 (secondary), and 0.42 (tertiary). Summary statistics of the resulting costs $\eta_{j,n}$ to the student are given in Appendix C.

3.4 Technology Parameters

The production technology parameters in the model include the share of labor α, the inverse elasticity of substitution μ, the N country-specific technology parameters I_n, and the diminishing return parameter ϕ.

The share of labor is set at 2/3, which is consistent with Gollin (2002)'s finding that capital's share ranges form 25 to 40 percent across countries and is uncorrelated with income levels.

[22] Here the numerator is the total government expenditures on education at level j per worker in the previous period. Dividing total government expenditures by F_n, the ratio of young generation agents to old generation agents, converts this into total government schooling expenditures per agent in the young generation. The total indirect cost of schooling at level j for agent θ is her wage $w(\theta)$ times the number of years of schooling at level j that agent θ attains. Integrating over all values of θ yields the average total cost of schooling at level j per agent in the young generation.

The elasticity of substitution in the model is the elasticity of substitution between *tasks*, but most estimates are of the elasticity of substitution between *workers* of different skill levels. For example, Katz and Murphy (1993) use a college/high school elasticity of 1.4, which would yield an inverse elasticity of 0.7. In this model, workers of different skill levels work in different tasks, but any worker can potentially perform any task and workers are perfect substitutes within tasks. In the context of the model, the elasticity of substitution between tasks may be less than the elasticity of substitution workers. I calibrate the baseline to $\mu = 0.7$, but check the robustness of the results to higher values (i.e., lower elasticities).

For each country n, I_n is calibrated to match average education levels.[23]

The parameter ϕ is calibrated to minimize the squared deviations of log output/worker between the model and the data. Formally, the calibrated ϕ is the nonlinear least squares estimate from:

$$\min_{\phi} var\left[\ln Y - \ln\left(K^{1-\alpha}L^{a}\right)\right] = \min_{\phi} var\left[\ln A\right]$$

That is, I choose ϕ to maximize the amount of output/worker variation that the model can explain and minimize the amount of variation in the residual technology parameter.

Values for $\underline{\theta}$, μ, ϕ, I_n, and A_n cannot be solved analytically, and are instead solved via simulation, which involves solving the relevant differential equations in the policy function. The details of the computation/simulation are outlined in Appendix D.

The calibrated technology values, together with the Mincerian returns, for the sample of countries are listed in Appendix E.

[23] If data on average years of schooling and GDP/worker are taken to be the true, the calibrated values correspond precisely to consistent point estimates from a method of moments estimation on a large, representative cross-section of workers.

4 Results

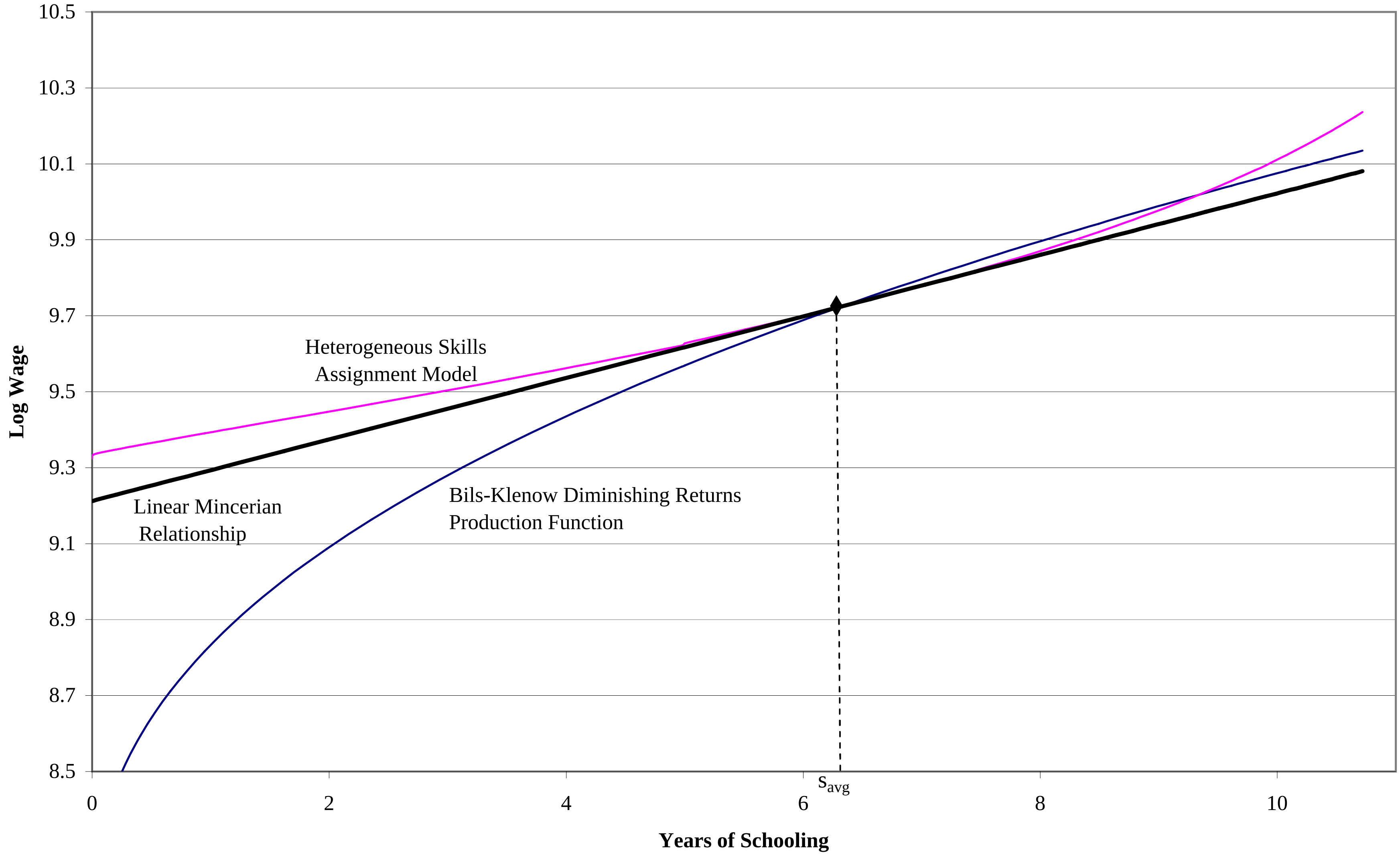
I evaluate, first, the model's success in producing the log-linear within country Mincerian relationship, the cross-country inverse relationship of Mincerian returns and schooling levels, and the fit of the model in predicting specific Mincerian returns. Next I discuss the counterfactual simulation results, then the model's implications for the true returns and growth effects of education, and finally robustness checks on the calibration and findings.

4.1 Baseline Model Fit

Figure 2 compares the log wage-schooling relationship in the model to the BK diminishing returns function and a linear Mincerian relationship. The specific country shown is Spain, which had the mean level of average years of schooling (6.3 years) and a Mincerian return (8.1 percent) quite close to the mean (8.8 percent). For illustration, the intercepts of the linear Mincerian relationship and BK functions have been adjusted to so that the functions are equal at the mean level of schooling. The BK diminishing returns relationship diverges sharply from the linear Mincerian relationship, but the model's log wage-schooling relationship is comparable to the linear Mincerian relationship, essentially linear except at extreme values.[24] Given the near-linear relationship, it is reasonable to run the linear Mincerian regression on the model's output to yield predicted Mincerian returns for each country. These predicted returns are negatively related to average school-

[24] The convexity at extreme levels of schooling is a result of the fixed reference point for discounting in τ. Given a fixed reference point (s_{avg}), additional years of schooling accumulate at the beginning of the worklife, which substantially shortens (discounted) working life. Thus, marginal returns increase noticably with years of schooling. In a sequential lifecycle problem, the reference point for discounting is the optimal schooling level, so discounted working time varies little with schooling and marginal returns are essentially constant.

Figure 2: Log Wage vs. Schooling Cross-Sectional Relationship (Spain*)



*Spain has the average years of schooling in the sample and a Mincerian return (0.081) close to the sample average (0.088).

ing in the cross-section of countries. The cross-country coefficient of -0.33 is comparable to the slope in the Figure 1 relationship of -0.30. Thus, the model does well in matching both the within-country log wage-schooling relationship and the cross-country Mincerian return-average schooling relationship.

At typical interest rates of 6.5 percent (implied by a 4 percent rate of time preference and 2.5 percent growth in consumption), the Mincerian returns in the model average just 6.0 percent, well below the 8.8 percent in the data. This is true despite the inclusion of direct costs of schooling, true[25] foregone earnings at all levels of schooling, and a large ability bias in the model. Alternative simulations showed that any additional costs (e.g. transportation, psychic costs) of schooling would have to average over three times foregone earnings to produce the level of Mincerian returns in the data. Such costs were viewed as unreasonable. Instead, as previously mentioned, higher interest rates (9.0 percent) were calibrated to reconcile the model with the data. The importance of this choice is discussed in the Robustness section below.

For this high interest baseline simulation, Figure 3 shows a scatterplot of predicted vs. actual Mincerian returns. The baseline summary statistics are given in the second column of Table 3. The returns in the model are significantly correlated (at a 1e-6 significance level) with those in the data with a correlation of 0.63 and explain a substantial fraction of the actual variation. To capture this, I construct a pseudo-R^2 value, as the analog of a regression R^2.[26] The pseudo-R^2 is bounded above at one, but can potentially be negative, especially since the model is calibrated to different moments and has

[25] In Mincer (1974), schooling does not subtract from the length of one's career, but only delays the career.

[26] Denoting x as the Mincerian return, and $\overline{x}$ as the average return, the formula is pseudo-$R^2 = 1 - \sum \left(x_i^{actual} - x_i^{predicted} \right)^2 / \sum \left(x_i^{actual} - \overline{x} \right)$

Figure 3: Mincerian Returns - Model vs. Data

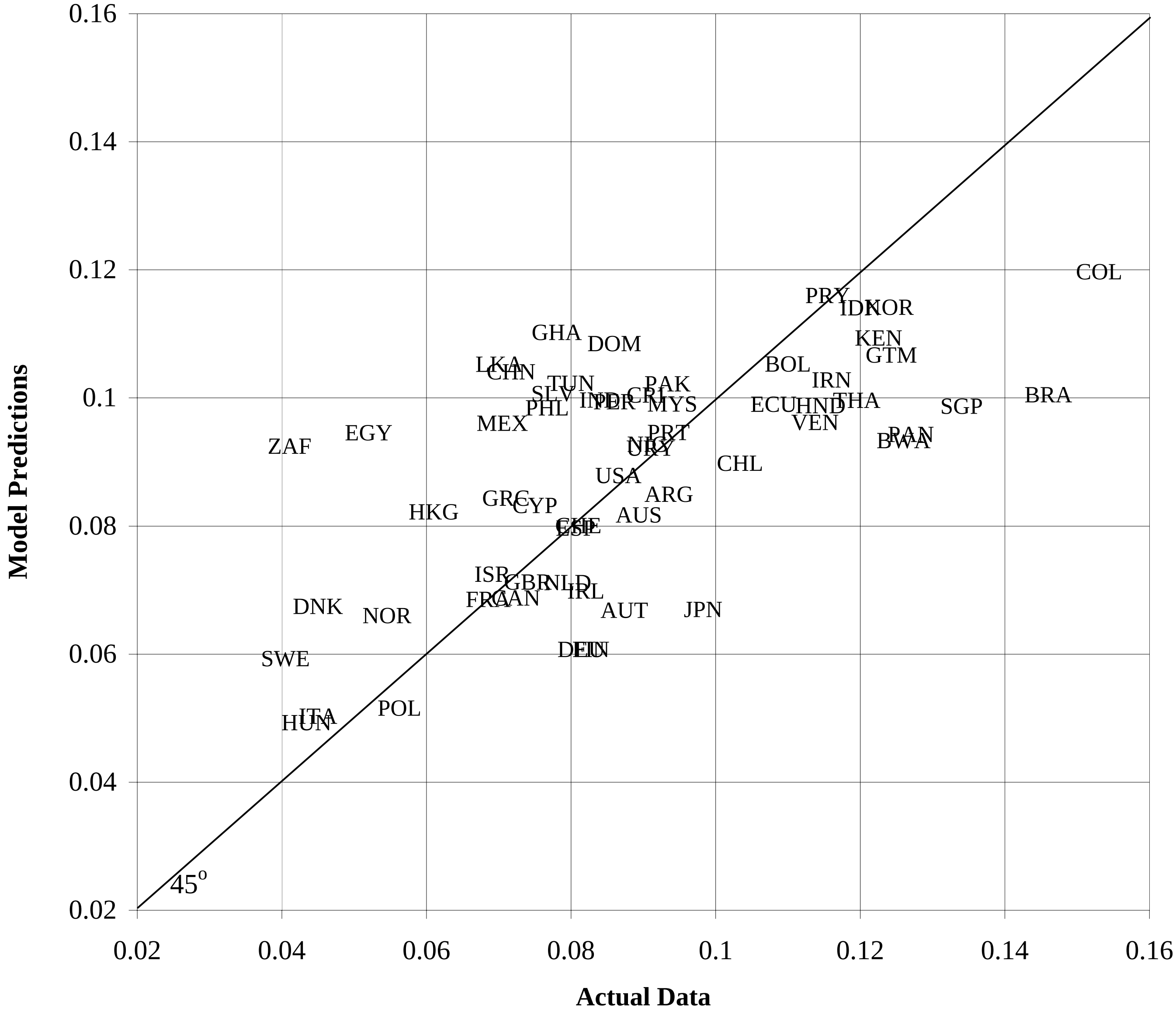


Table 3: Simulations Summary Statistics

Variable	Actual Data	Baseline Model	Counterfactual Simulations: No Variation	Only Skill-Biased Technology (I) Variation	Only Career Length (T) Variation	Only Discount Rate (r) Variation	Direct Costs: Only Non-Fertility Direct Cost (η*F) Variation	Only Fertility (F) Variation
Mincerian Returns								
Average	8.8%	8.8%	8.8%	8.8%	8.8%	8.8%	8.8%	8.8%
Standard Deviation	2.6%	1.8%	0.0%	0.5%	0.2%	0.4%	0.5%	0.5%
Pseudo-R^2	--	0.40	0.00	0.01	0.05	0.03	0.19	0.16
Correlation with Actual	1.00	0.63	0.00	0.13	0.40	0.17	0.57	0.50
Years of Schooling								
Average	6.2	6.2	6.2	5.7	6.3	6.2	6.2	6.2
Standard Deviation	2.5	2.5	0.0	1.5	0.2	0.4	0.7	0.6
Pseudo-R^2	--	1.00	0.00	0.29	0.09	-0.06	0.07	0.30
Correlation with Actual	1.00	1.00	0.00	0.58	0.72	-0.16	0.27	0.72
Log Income/Capita								
Average	9.38	9.38	9.38	9.46	9.64	9.62	9.61	9.62
Standard Deviation	0.79	0.79	0	0.6	0.008	0.02	0.03	0.03
Pseudo-R^2	--	1.00	0.00	-0.09	-0.10	-0.10	-0.07	-0.05
Correlation with Actual	1.00	1.00	0.00	0.33	0.78	-0.15	0.34	0.66

Notes: The baseline model matches average years of schooling and income/capita country by country in the data by construction. The constant parameters in the no variation counterfactual were chosen individually to match the average Mincerian returns in the cross-section of countries to the baseline model. The constant skill-neutral technology parameter (A) was chosen to match average log income levels in the sample. "Pseudo-R^2" of variable X is $1-(\Sigma(X_{data}-X_{model})^2/\Sigma(X_{data}-X_{mean})^2$.

no free parameters. Still the model produces an R^2 of 0.40. For comparison, linear regression models, which choose free parameters to maximize R^2, have much less explanatory power. The Bils-Klenow regression of log Mincerian returns on log schooling levels has one degree of freedom but yields an R^2 of just 0.16[27], while linear regression of the Mincerian returns on all the relevant cross-country variables (i.e., career length T, discount rate r, fertility F, governmnent funding levels e_1, e_2 and e_3, and average schooling level) raises this R^2 to just 0.32.

A simple analysis of the errors in Figure 3 uncovers no glaring cause of over- or under-prediction. The errors are not significantly or even sizably correlated with average schooling levels or any of the key data variables listed above in the model. Neither are they significantly or sizably correlated with any of the Barro-Lee school quality measures. This indicates that not accounting for differences in quality is probably not a major problem. Two variables were found to be correlated with the errors, though neither can explain more than five percent of the variation in actual returns. First, countries in Africa tend to have returns about 0.02 points lower than the model predicts. (From Figure 3, one can see that South Africa, Egypt and Ghana have substantially lower returns than the model predicts, while Botswana has a substantially higher return). Second, countries with larger (log) capital stocks have substantially higher returns than the model predicts. Doubling the capital stock would increase the difference between actual and predicted returns by 0.005 points. A theory of capital-skill complementarity might contribute to increased demand for educated labor in high capital countries, but the theory would also need to explain how capital increases the costs of schooling. In addition, there is a significant negative relationship between

[27] The pseudo-R^2 of the model is 0.40 in terms of explaining both the Mincerian return and the log Mincerian return.

the *absolute value* of the error terms and log output per worker or log capital stock (R^2 of 0.07 and 0.09, respectively). Since data quality generally improves with level of development, some of the discrepancy between model and data is likely caused by measurement error in either the explanatory data or the Mincerian return data.)

4.2 Counterfactuals

Counterfactuals simulations isolate the roles of different sources of variation in the simulated distribution of Mincerian returns, schooling levels, and income per capita. These counterfactuals were simulated using variation in only a single factor (e.g. fertility, career lengths), while "turning off" (i.e. equalizing[28]) all other factors across countries, in order to see the effect on the distributions of Mincerian returns, schooling levels, and income per capita across countries. Two types of counterfactuals were used that affect η_j, the ratio of indirect to direct schooling costs. The first of these counterfactuals is the effect of fertility variation. The second counterfactual is the effect of variation in $\eta_j * F_j$, or direct costs variation not related to fertility. This second counterfactual is driven by variation in government expenditures relative to enrollment rates. For both direct cost counterfactuals, I report the total effect of variation at the primary, secondary, and tertiary levels.

Returning attention to Table 3, one can compare the counterfactual results to both the no variation and the baseline results. In explaining the

[28] For a given factor Z, the "equalized" constant value, $\overline{Z}$, is the value that matches the average Mincerian return across all 59 countries, when Z is the only factor equalized. A and K do not affect Mincerian returns, however, so $\overline{K}$ is cross-sectional average of K_n and $\overline{A}$ matches average output per capita in the cross-section of countries. The simulations in which variation in a single factor was "turned off" relative to the baseline produced the same ordering of the importance of the different factors.

variation in Mincerian returns, variation in the direct costs of schooling (caused by variation in government expenditures and fertility) is most important. Within this category, non-fertility variation produces a pseudo-R^2 of 0.19, while the fertility variation alone produces a pseudo-R^2 of 0.16.

Compared to these school funding variables, career length and discount rate variation play relatively small roles. The discount rate has a small effect because it does not vary much in the data (see Table 1). Career length plays a relatively small role in part because of the high calibrated discount rate needed to match average returns. After discounting, additional years at the end of a career have only small effects on lifetime earnings.

Skill-biased technology explains very little of the Mincerian return variation, but plays a very important role in explaining output per capita and schooling attainment. Though the R^2 are less meaningful for schooling and output (because the means in the counterfactuals deviate from the means in the data), the large standard deviation in the counterfactual results and correlation of the counterfactual results with the data support this claim.

In summary, supply factors affecting the direct costs of schooling play the most important role in explaining the returns to schooling, while demand factors (i.e. skill-biased technology) affects the level of schooling and output.

4.3 Schooling and Output

Although Mincerian returns in the model average 8.8 percent, the true returns to additional years of schooling are much lower. The reason for this can be seen by examining a heuristic version of equation (6), the household's first-order condition for choosing schooling:

$$\frac{\partial \ln w(i,h)}{\partial h} = \frac{(w + \eta w)}{(\tau w - \eta s w)}$$

In words, the ratio of total (indirect and direct) marginal costs of an additional schooling to lifetime income net of direct schooling costs must equal the marginal wage return to schooling. Direct costs of schooling are small relative to foregone earnings (i.e., η is highest at the tertiary level and still averages just 0.18), and so the total cost of a year of schooling is small relative to discounted lifetime earnings net of schooling costs.[29] This holds despite the additional 2.7 percent added to interest rates in order to match the high returns in the data.

Given diminishing returns, incremental discrete gains in schooling are even smaller than marginal returns. The assumption of log diminishing returns is required for returns to schooling decrease as levels of schooling increase, which is observed empirically. In Table 4 shows that discrete percent wage returns:

$$\frac{E\left[w(i,h)-w(i,h+n)\right]}{E[w(i,h)]}$$

for $n = 1, 2$, and 4 years of schooling average six percent or less.

The presence of diminishing returns makes the output effects from additional years of schooling (given constant technology) smaller than marginal effects would predict, but also indicates that existing investments have had larger growth effects. Table 5 presents results for simulations that set schooling levels to zero for all people in all countries. Output per worker is on average 31 percent lower in the "no schooling" simulations, with the U.S. having the largest decline in output (66 percent). Restated, if the U.S. workforce had no education, the model predicts output/worker would be only one-third of actual output/worker. The average percent wage returns for existing schooling average 6.4 percent and are again somewhat higher than the returns on discrete increases in Table 4. The U.S. has among the largest average returns for existing schooling at 9.4 percent/year.

[29] Realizing this issue, BK add utility costs to schooling in order to reconcile observed Mincerian returns and observed schooling levels within a representative agent model.

Table 4: Diminishing Returns: Average Returns per Year to Discrete Changes in Schooling

	Average Gain Per Year in Model		
Variable	**1-Year Gain**	**2-Year Gain**	**4-Year Gain**
Mean	6.0%	5.8%	5.4%
Standard Deviation	1.2%	1.2%	1.1%
Correlation with Mincerian Return in Model	0.66	0.63	0.58
Correlation with Mincerian Return in Data	0.44	0.42	0.39

Table 5: Effect of No Education on Output per Worker

Code	Country	Percentage Decrease in Output/Worker	Average Return/Year	Code	Country	Percentage Decrease in Output/Worker	Average Return/Year
ARG	Argentina	16%	4.9%	ITA	Italy	38%	4.3%
AUS	Australia	53%	7.9%	JPN	Japan	44%	6.4%
AUT	Austria	32%	5.3%	KEN	Kenya	14%	5.7%
BOL	Bolivia	23%	6.6%	MYS	Malaysia	32%	7.2%
BWA	Botswana	11%	4.6%	MEX	Mexico	34%	7.2%
BRA	Brazil	18%	5.6%	NLD	Netherlands	38%	5.7%
CAN	Canada	45%	6.0%	NIC	Nicaragua	16%	5.4%
CHL	Chile	34%	7.0%	NOR	Norway	34%	5.4%
CHN	China	31%	7.4%	PAK	Pakistan	10%	4.7%
COL	Colombia	27%	7.8%	PAN	Panama	44%	8.0%
CRI	Costa Rica	32%	7.3%	PRY	Paraguay	31%	8.1%
CYP	Cyprus	41%	7.0%	PER	Peru	20%	6.0%
DNK	Denmark	26%	4.9%	PHL	Philippines	41%	8.1%
DOM	Dominican Rep.	34%	7.9%	POL	Poland	32%	4.1%
ECU	Ecuador	33%	7.4%	PRT	Portugal	46%	8.3%
EGY	Egypt	18%	5.6%	ZAF	S. Africa	25%	6.2%
SLV	El Salvador	17%	5.8%	KOR	S. Korea	61%	10.9%
FIN	Finland	41%	5.6%	SGP	Singapore	32%	7.2%
FRA	France	29%	5.2%	ESP	Spain	30%	6.0%
GHA	Ghana	14%	5.5%	LKA	Sri Lanka	33%	7.7%
GRC	Greece	41%	7.1%	SWE	Sweden	35%	4.7%
GTM	Guatemala	12%	5.0%	CHE	Switzerland	45%	7.0%
HND	Honduras	19%	6.0%	THA	Thailand	30%	7.2%
HKG	Hong Kong	45%	7.3%	TUN	Tunisia	14%	5.3%
HUN	Hungary	26%	3.6%	GBR	U.K.	38%	5.7%
IND	India	18%	5.9%	USA	U.S.	66%	9.4%
IDN	Indonesia	25%	7.4%	URY	Uruguay	38%	7.3%
IRN	Iran	17%	5.9%	VEN	Venezuela	27%	6.7%
IRL	Ireland	39%	6.2%	DEU	W. Germany	32%	4.4%
ISR	Israel	44%	6.6%		**Average**	**31%**	**6.4%**

4.4 Robustness

The Mincerian return predictions are remarkably robust. The predicted effects of schooling on output/worker are smallest for the calibrated simulation, and so are viewed as conservative estimates. A summary of these results:

- The Mincerian return results were quite robust to higher values (0.99, 0.9, and 0.8) of μ. However, the effects of schooling on output per worker were stronger for higher μ.

- Simulations at lower real interest rates (6.5 percent) but with an additional cost of schooling calibrated to match the average Mincerian return in the data left the pseudo-R^2 and effect of schooling on output per worker virtually unchanged. With less discounting, career length explains somewhat more of the variation in the returns (the pseudo-R^2 for career length variation alone was 0.10 instead of 0.05), but still less than direct cost variation.

- The results on observed Mincerian returns were not substantially effected by the choice of ϕ. Within a range of 0.6 to 0.8, the output/worker loss in the no schooling predictions were larger but averaged less than 50 percent. For $\phi > 0.8$ the model could not be calibrated to match the interquartile range of schooling, while for values below 0.6 the predicted impact of schooling per worker increased substantially.

- The ability bias in the model was lowered by decreasing the ability spread $\overline{\theta} - \underline{\theta}$ (for various values of ϕ) to match the cross-country relationship between Mincerian returns and interquartile schooling . These simulations required even larger real interest rates to match average returns, but at these rates the R^2 remained near 0.40. The effects of schooling on output per worker were again larger in these models, averaging as high as 85 percent of output.

- Bounding subsidies to schooling at a maximum 33% of the indirect costs had almost no effect on the results.

5 Conclusions

A model with simple supply and demand variation, but no major institutional differences (e.g. differences in labor markets/institutions, credit markets, inequality, or school quality) yielded reasonable simulated results able to explain about forty percent of the variation in the data, but required high real interest to explain the high level of returns. Variation in the direct costs of schooling resulting from different fertility rates and funding rates contributed the most to this variation in the data. Despite smaller than measured true returns, schooling has sizable effects on output levels.

The paper also uncovered several questions for future research. First, can any observable or measurable variables explain the deviations of predicted returns from true returns? Second, can observable variables explain the high level of Mincerian returns?

References

[1] Acemoglu, D. "Technical Change, Inequality, and the Labor Market." NBER working paper 7800, July 2000

[2] Acemoglu, D. "Directed Technical Change." NBER working paper 8287, May 2001

[3] Bils, M. and Klenow, P. J. "Does Schooling Cause Growth?" *American Economic Review.* 90 (December 2000): 1160-83

[4] Berman, E., Bound, J., and Griliches, Z. "Changes in the Demand for Skilled Labor within U. S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturers." *Quarterly Journal of Economics,* 109 (May 1994): 367-397

[5] Borghans, L. and ter Weel, B. "The Diffusion of Computers and the Distribution of Wages." mimeo, 2002

[6] Caselli, F. "Technological Revolutions." *American Economic Review,* 89 (March 1999): 78-92

[7] Cawley, J., Heckman, J., and Vytlacil, E. "Meritocracy in America: Wages Within and Across Occupations." *Industrial Relations*, 38 (July 1999): 250-296

[8] Goldin, C. and Katz, L. "The Returns to Skill in the United States Across the Twentieth Century." NBER working paper W7126, May 1999

[9] Goldin, C. and Katz, L. "The Origins of Capital-Skill Complementarity." *Quarterly Journal of Economics,* 113 (August 1998): 693-732

[10] Goldin, C. and Katz, L. "The Decline of Non-Competing Groups: Changes in the Premium to Education, 1890 to 1940." NBER working paper 5202, 1995

[11] Greenwood, J. and Yorukoglu, M. "1974" *Carnegie-Rochester Conference Series on Public Policy,* XLVI (1997): 363-382

[12] Hall, R. and Jones, C. "Why Do Some Countries Produce So Much More Output per Worker than Others?" *Quarterly Journal of Economics*, 114 (February 1999): 83-116.

[13] Heckman, J. and Klenow, P. "Human Capital Policy." mimeo, 1997

[14] Heckman, J., Lochner, L. and Taber, C. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics,* 1 (January 1998): 1-58

[15] Heckman, J. and Vytlacil, E. "Policy-Relevant Treatment Effects." *American Economic Review, 91* (May 2001):107-112

[16] Kaboski, J. "Education, Sectoral Composition, and Growth." mimeo, 2005

[17] Katz, L. and Murphy, K. M. "Changes in the Wage Structure 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics,* 107 (February 1992): 35-78

[18] Klenow, P. J. and Rodigruez-Clare, A. "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?" in Bernanke, B. S. and Rotemberg, J. J. (ed.), *NBER Macroeconomics Annual 1997.* Cambridge, MA: MIT Press, 1997

[19] Krueger, A. "How Computers Have Changed the Wage Structure: Evidence from Micro-Data, 1984-1989." *Quarterly Journal of Economics,* 108 (February 1993): 33-60

[20] Krueger, A. and Lindahl, M. "Education for Growth: Why and for Whom?" *Journal of Economic Literature,* 39 (December 2001): 1101-1136

[21] Krusell, P., Ohanian, L. E., Rios-Rull, J. V., and Violante, G. L. “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis.” *Econometrica*, 68 (September 2000): 1029-53

[22] Lloyd-Ellis, H. “Endogenous Technological Change and Wage Inequality.” *American Economic Review,* 89 (March 1999): 44-77

[23] Machin, S. and Van Reenen, J. “Technology and Changes in Skill Structure: Evidence from Seven OECD Countries.” *Quarterly Journal of Economics,* 113 (November 1998): 1215-1244

[24] Mincer, J. *Schooling, Experience and Earnings.* New York: Columbia University Press, 1974

[25] Murphy, K. M. and Welch, F. “Occupational Change and the Demand for Skill, 1940-1990.” *American Economic Review,* 83 (May 1993): 122-126

[26] Psacharopoulos, G. and Patrinos, H. “Returns to Investment in Education: A Further Update." World Bank Policy Research Working Paper 2881, September 2002

[27] Romer, P. “Endogenous Technological Change.” *Journal of Political Economy,* 98 (October 1990): S71-S102

[28] Sattinger, M. “Comparative Advantage and the Distribution of Earnings and Abilities.” *Econometrica,* 43 (May 1975): 455-468

[29] Teulings, C. N. “The Wage Distribution in a Model of the Assignment of Skills to Jobs.” *Journal of Political Economy,* 103 (April 1995): 280-313

[30] Tinbergen, J. *Income Distribution: Analysis and Policies,* Amsterdam: North-Holland, 1975

[31] Winter-Ebmer, R. "Public Funding and Enrollment into Higher Education in Europe", IZA Discussion Series Paper #503, 2002

[32] "Rates of Return to Investment in Education." World Bank Edstats dataset, 2002

[33] World Development Indicators 2002, cd-rom, 2002

6 Appendix A

The necessary[30] first-order conditions for the household's choice of schooling is:

$$
\begin{array}{ll}
\frac{\partial \ln w(i,h)}{\partial h} \leq \frac{\tau'+\eta_1}{\tau(0;T,r)} & \text{for } s=0 \\
\frac{\partial \ln w(i,h)}{\partial h} = \frac{\tau'+\eta_1}{\tau(s;T,r)-\eta_1 s} & \text{for } 0<s<\bar{s}_1 \\
\frac{\partial \ln w(i,h)}{\partial h} \in \left(\frac{\tau'+\eta_1}{\tau(\bar{s}_1;T,r)-\eta_1\bar{s}_1}, \frac{\tau'+\eta_2}{\tau(\bar{s}_1;T,r)-\eta_1\bar{s}_1}\right) & \text{for } s=\bar{s}_1 \\
\frac{\partial \ln w(i,h)}{\partial h} = \frac{\tau'+\eta_2}{\tau(s;T,r)-[(\eta_1-\eta_2)\bar{s}_1+\eta_2 s]} & \text{for } \bar{s}_1<s<\bar{s}_2 \\
\frac{\partial \ln w(i,h)}{\partial h} \in \left(\frac{\tau'+\eta_2}{\tau(\bar{s}_2;T,r)-[(\eta_1-\eta_2)\bar{s}_1+\eta_2\bar{s}_2]}, \frac{\tau'+\eta_3}{\tau(\bar{s}_2;T,r)-[(\eta_1-\eta_2)\bar{s}_1+\eta_2\bar{s}_2]}\right) & \text{for } s=\bar{s}_2 \\
\frac{\partial \ln w(i,h)}{\partial h} = \frac{\tau'+\eta_3}{\tau(s;T,r)-[(\eta_1-\eta_3)\bar{s}_1+(\eta_2-\eta_3)(\bar{s}_2-\bar{s}_1)+\eta_3 s]} & \text{for } \bar{s}_2<s<\bar{s}_3 \\
\frac{\partial \ln w(i,h)}{\partial h} \geq \frac{\tau'+\eta_3}{\tau(\bar{s}_3;T,r)-[(\eta_1-\eta_3)\bar{s}_1+(\eta_2-\eta_3)(\bar{s}_2-\bar{s}_1)+\eta_3\bar{s}_3]} & \text{for } s=\bar{s}_3
\end{array}
\tag{15}
$$

[30] In the case where $\eta_1 \leq \eta_2 \leq \eta_3$, the problem is concave in s and these conditions are sufficient conditions. Otherwise, these conditions are only necessary conditions, and it is possible that the above conditions hold for multiple values of s. The maximum can be found by comparing the objective function across the finite number of potential s values satisfying (15).

7 Appendix B

7.1 Imputed Data

The three modification or imputations that were made involved: 1) the government educational expenditure data, 2) the fertility and life expectancy data, and 3) indicators for countries with fully funded primary, secondary, and/or tertiary education.

1. The government expenditures at a given level of education as a fraction of output (e_j) are calculated from data on expenditures per pupil by level of education, enrollments at different levels, and the government educational expenditures as a fraction of GDP. That is, for each level of education j, I multiply the expenditures per pupil at level j by enrollment at level j, to yield total expenditures at level j. I then use these values of total expenditures to get the relative split of total expenditures between primary, secondary, and tertiary education. I then multiply these by educational expenditures as a fraction of GDP, to yield e_j values.[31] Finally, higher education expenditures were adjusted downward to account for the fact much of higher education expenditures are for research services, not educational services. The downward adjustment factor of 0.5 is consistent with available data on the relative fraction of expenditures used for education and research.[32]

[31] For two countries, West Germany and Indonesia, I had missing data and could only get the relative split of expenditures between primary and secondary levels. For these values, I use the sample average to get the split between between tertiary and non-tertiary spending, then used the primary-secondary split to subdivide non-tertiary expenditures.

[32] Available data (NCES, 2001) are for the U.S. in 1996-97: obvious educational services (instruction and student services) totaled $72.3 billion, while obvious non-educational services (i.e. research, hospital services and independent operations, auxiliary services, and public services) totaled $68.1 billion. Other services that are less easily categorized

2. Since the choice between work and education in the model begins at the primary school entrance age, I adjust fertility (downward) and life expectancy (upward) to eliminate variation associated with infant mortality variation. Thus, I use total fertility of children that survive beyond the first year, and life expectancy of children that survive beyond the first year.

3. Countries with compulsory primary and/or secondary education were assumed to fully fund those levels of education, respectively. The list countries with fully funded tertiary education was based on meeting at least one of two criteria: 1) over 90 percent of tertiary expenditures were public based in 1995 or 1999 based on an OECD data of all OECD and some non-OECD countries (OECD, 2002), and 2) having no tuition costs for European countries (Winter-Ebmer, 2002). This data on which countries have fully funded education is only used to calibrate the cost of schooling parameters η_j, as discussed in Section 3.2.2, and so need not be a complete.

7.2 Timing

The exercise aims to match the international cross-section of economies in 1990, which was chosen because it was the year of best data availability. A summary of other timing decisions in the data:

- 1990 values were used for capital/worker and educational attainment distributions. Since output per worker is a flow with cyclical variation, I used average output per worker from 1988-1992.

(i.e. libraries and other academic services, plant and operations, and institutional support) totaled \$31.1 billion.

- For fertility, life expectancy, earlier years are used, namely the averages for the years 1960, 1965 and 1970.

- For the duration of primary and secondary schooling, the averages for the years 1965, 1970 and 1975 are used. These data are relatively stable over time.

- For retirement ages, the available data is for 1990 and 1999. Since, I am modeling the labor force that is working in 1990, I use the 1999 values.

- Data on educational expenditures is not available for all countries in earlier years. Therefore, in calculating subsidies, I use time averages of the fraction of government educational expenditures as a share of output, as well as the relative distribution of these expenditures across primary, secondary and tertiary levels. Total government expenditures as a fraction of output do not show a strong trends within countries during this period. The distributions of these trends across primary, secondary and tertiary education do exhibit small trends whose direction varies from country to country. Estimating either a global trend or country-specific trends is problematic because of missing observations and compositional changes in the sample. In the case of both total government expenditures and the distribution of expenditures across schooling levels, the time variation within a country is small compared to the cross-country variation, however.

- The years of available Mincerian return estimates is much more sporadic and country-specific. Again, the time variation is small-compared to the cross-country variation, so I again use time averages of the available data.

- The only available data on years of compulsory schooling and the age of entrance into primary school were for 1993 and 1997 . If available, I used the 1993 value. Otherwise, the 1997 value was used.

7.3 Calibration of Schooling Costs

I calculate the expectation in (14) by transforming this equation into an equivalent integral in s :

$$\tilde{\eta}_{j,gov} = \frac{e_{j,-1} Y_{-1}}{F \int s_j w(s) \upsilon(s) ds}$$

The distribution of schooling levels in the data is used to form a density of schooling levels $\upsilon(s)$ and s_j is easily calculated directly from schooling duration and attainment data. Given the Cobb-Douglas technology, wages are proportional to income per worker, and so the $\frac{w(s)}{Y}$ relationship is approximated using the Mincerian return data.

That is, given the Mincerian return m, wages can be expressed:

$$w(s) = w(0) e^{ms}$$

where $w(0)$ is an unknown constant. Total discounted wage earnings are:

$$\text{discounted average lifetime wage earnings} = \int_0^{16} \tau(s) w(0) e^{ms} \upsilon(s) ds$$

In the data (the source of e) educational expenditures are paid across cohorts and over time, income should not be discounted since educational expenditures are not. So I multiply discounted earnings by the ratio of average real time to average discounted time $(T - s_{avg}) / \tau(s_{avg})$ to yield average lifetime wage earnings:

$$\text{avg. lifetime wage earnings} = (T - s_{avg}) \int_0^{16} \frac{\tau(s)}{\tau(s_{avg})} w(0) e^{ms} \upsilon(s) ds$$

Given the Cobb-Douglas technology, labor is a constant share so $\frac{w(0)}{Y}$ can be solved using the following formula::

$$\begin{aligned} (T - s_{avg}) \int_0^{16} \frac{\tau(s)}{\tau(s_{avg})} w(0) e^{ms} \upsilon(s) ds &= \alpha Y \\ \frac{\alpha}{(T - s_{avg}) \int_0^{16} \frac{\tau(s)}{\tau(s_{avg})} e^{ms} \upsilon(s) ds} &= \frac{w(0)}{Y} \end{aligned}$$

The formula for calibrating $\tilde{\eta}_j$ is:

$$\tilde{\eta}_j = \frac{\sum_n FF_{j,n} \tilde{\eta}_{j,gov,n}}{\sum_n FF_{j,n}}$$

Given the Cobb-Douglas technology, labor is a constant share so $\frac{w(0)}{Y}$ can be solved using the following formula:

$$\begin{aligned} \int_0^{16} \tau(s) w(0) e^{ms} \upsilon(s) ds &= \alpha Y \\ \frac{\alpha}{\int_0^{16} \tau(s) e^{ms} \upsilon(s)} &= \frac{w(0)}{Y} \end{aligned}$$

I calculate $\tilde{\eta}_{j,gov}$ for each country and given $\tilde{\eta}_{j,gov,n}$ the formula for calibrating $\tilde{\eta}_j$ is then:

$$\tilde{\eta}_j = \frac{\sum_n FF_{j,n} \tilde{\eta}_{j,gov,n}}{\sum_n FF_{j,n}}$$

where $FF_{j,n}$ represents the country indicator variable for fully funded education at level j.

8 Appendix D

The actual ODE solved is not (11), but its equivalent ODE in terms of the inverse $i(\theta)$. Using the chain rule, I first substitute $h'(i) = h'(\theta)\theta'(i$ into (11):

$$\frac{\theta''}{\theta'}+\left(\frac{g'}{g}-\frac{\tau'}{\tau\left(h-\theta;T,r\right)}\right)\theta'+\left(\frac{a_2}{a}+\frac{\tau'}{\tau\left(h-\theta;T,r\right)}\right)h'(\theta)\theta'+\left(\frac{\mu-1}{\mu}\right)\frac{a_1}{a}=0 \tag{16}$$

As noted in the paper, the form of $\tau\ (s;T,r,s_{avg})$ is a linear approximation of $\tilde{\tau}$ in equation (13):

$$\begin{aligned}\tau(s;T,r,s_{avg}) &= c_1+c_2 s\\ c_1 &\equiv \frac{e^{-r(T-s_{avg})}-1}{-r}+s_{avg}\\ c_2 &\equiv -1\end{aligned}$$

Substituting in for τ,using the inverse rule for derivatives and rearranging, we can write:

$$i''(\theta)=\left(\frac{g'(\theta)}{g}+\frac{c_2\left[h'(\theta)-1\right]}{c_1+c_2(\theta-h)}+\frac{a_2}{a}h'(\theta)\right)i'(\theta)+\left(\frac{\mu-1}{\mu}\right)\frac{a_1}{a}\left[i'(\theta)\right]^2 \tag{17}$$

where we have again omitted the functional dependency of $g, a,$ and h.

Given the uniform distribution for θ, $\frac{g'}{g}=0$. Differentiation of a yields:

$$\begin{aligned}a(i,h) &= \exp\left(\frac{i^{1+\phi}h^{1-\phi}}{(1+\phi)(1-\phi)}\right)\\ \frac{a_1}{a}(i,h) &= \frac{i^{\phi}h^{1-\phi}}{(1-\phi)}\\ \frac{a_2}{a}(i,h) &= \frac{1}{(1+\phi)}\frac{i^{1+\phi}}{h^{\phi}}\end{aligned}$$

Finally, h is solved numerically using the optimality conditions for s

(equation (15)):

$$
\begin{array}{ll}
h=\theta & \text{for } s=0 \\
\frac{1}{(1+\phi)}\frac{i^{1+\phi}}{h^{\phi}}=\frac{-c_2+\eta_1}{c_1+c_2(h-\theta)} & \text{for } 0<s<\bar{s}_1 \\
h=\theta+\bar{s}_1 & \text{for } s=\bar{s}_1 \\
\frac{1}{(1+\phi)}\frac{i^{1+\phi}}{h^{\phi}}=\frac{-c_2+\eta_2}{c_1-(\eta_1-\eta_2)\bar{s}_1+(c_2-\eta_2)(h-\theta)} & \text{for } \bar{s}_1<s<\bar{s}_2 \\
h=\theta+\bar{s}_2 & \text{for } s=\bar{s}_2 \\
\frac{1}{(1+\phi)}\frac{i^{1+\phi}}{h^{\phi}}=\frac{-c_2+\eta_3}{c_1-(\eta_1-\eta_3)\bar{s}_1-(\eta_2-\eta_3)(\bar{s}_2-\bar{s}_1)+(c_2-\eta_3)(h-\theta)} & \text{for } \bar{s}_2<s<\bar{s}_3 \\
h=\theta+\bar{s}_3 & \text{for } s=\bar{s}_3
\end{array}
\tag{18}
$$

The implicit function theorem yields $h'(\theta)$:

$$
h'(\theta)=\begin{cases}
1 \\
\qquad \text{for } s=0,\bar{s}_1,\bar{s}_2,\bar{s}_3 \\
\dfrac{\left[\frac{(c_2-\eta_1)^2}{[c_1+c_2(h-\theta)]^2}-i'\left(\frac{i}{h}\right)^{\phi}\right]}{\left[\frac{c_2(1+\eta_1)}{[c_1+c_2(h-\theta)]^2}-\frac{\phi}{1+\phi}\left(\frac{i}{h}\right)^{1+\phi}\right]} \\
\qquad \text{for } 0<s<\bar{s}_1 \\
\dfrac{\left[\frac{(c_2-\eta_1)^2}{[c_1-(\eta_1-\eta_2)\bar{s}_1+(c_2-\eta_2)(h-\theta)]^2}-i'\left(\frac{i}{h}\right)^{\phi}\right]}{\left[\frac{c_2(1+\eta_1)}{[c_1-(\eta_1-\eta_2)\bar{s}_1+(c_2-\eta_2)(h-\theta)]^2}-\frac{\phi}{1+\phi}\left(\frac{i}{h}\right)^{1+\phi}\right]} \\
\qquad \text{for } \bar{s}_1<s<\bar{s}_2 \\
\dfrac{\left[\frac{(c_2-\eta_1)^2}{[c_1-(\eta_1-\eta_3)\bar{s}_1-(\eta_2-\eta_3)(\bar{s}_2-\bar{s}_1)+(c_2-\eta_3)(h-\theta)]^2}-i'\left(\frac{i}{h}\right)^{\phi}\right]}{\left[\frac{c_2(1+\eta_1)}{[c_1-(\eta_1-\eta_3)\bar{s}_1-(\eta_2-\eta_3)(\bar{s}_2-\bar{s}_1)+(c_2-\eta_3)(h-\theta)]^2}-\frac{\phi}{1+\phi}\left(\frac{i}{h}\right)^{1+\phi}\right]} \\
\qquad \text{for } \bar{s}_2<s<\bar{s}_3
\end{cases}
\tag{19}
$$

After substituting, the ODE (17) is solved using Matlab's ode15s function and a shooting algorithm to solve the boundary conditions: $i(\underline{\theta})=0$ and $i(\overline{\theta})=I$.