Labor Substitutability among Schooling Groups

Mark Bils
U. of Rochester, NBER

Baris Kaymak
Université de Montréal, CIREQ

Kai-Jie Wu
University of Rochester

July 6, 2020

Abstract

The degree of substitutability between schooling groups is essential to understanding the role of human capital in income differences and for assessing the economic impact of such policies as schooling subsidies, redistributive taxes, or selective immigration policies. We marshal information on educational attainments and wage premia for schooling (Mincer return) across countries and over time to identify the likely magnitude of substitutability among schooling groups. That magnitude is captured by the extent the Mincer return depends on relative scarcity of more educated workers, in addition to the quality of schooling and the skill bias in technology. For plausible patterns in schooling quality and technological change, we place the long-run elasticity of substitution on the order of 4, which is far higher than that commonly used in the literature.

1 Introduction

Returns to human capital investment, especially returns to schooling, play a prominent role in explanations of income differences within countries, differences across countries, and advancements in income over time. Mincerian returns to schooling, the gradient of wages with respect to years of schooling, has been a key tool for gauging returns to schooling investments. But this Mincerian return is, at best, a measure of private returns to schooling. Jumping to implications for cross-country income differences (levels accounting) or growth accounting requires some springboard of assumptions. As discussed in Jones (2014), Caselli and Ciccone (2019), and

Hendricks and Schoellman (2019), the mapping from Mincerian returns to human capital accumulation depends critically on how substitutable are workers with differing levels of schooling. If substitutability is low, then Mincerian returns largely reflect relative scarcities of workers with more versus less schooling. In turn, how Mincerian returns vary across countries or over time may largely reflect patterns in the relative scarcities.

Knowing the substitutability across groups is important, not only for levels and growth accounting, but also for judging the impact on inequality of policies to subsidize education. Tinbergen (1975) analyzed how expanding the supply of more-educated workers, under imperfect substitution, would drive down the return to schooling thereby offsetting upward trends in inequality from skill-biased technological change.[1] Substitutability across skills has a role more generally in analyzing policies or events that shift labor supply by skill. For instance, if substitutability is low, a policy to redistribute income that reduces the relative employment or hours of more-skilled workers will create an offsetting increase in inequality by driving up returns to skill, with the converse true if the policy especially reduces hours for less-skilled workers (e.g., Feldstein, 1973). Similarly, assessing the wage impact of selective immigration policies or of mass migration events in times of geopolitical distress, because they shift relative employments by skill, require knowledge of that substitutability.

The consensus in the literature is that the elasticity of substitution across workers with differing levels of schooling is quite low.[2] This consensus largely reflects works by Katz and Murphy (1992), Heckman, Lochner, and Taber (1998), and Card and Lemieux (2001), all of whom estimate an elasticity of labor demand between high school and college-trained workers for the U.S. of about 1.5.[3] Each estimates that elasticity first controlling for longer-term trends in relative wages. (Long-term trends have typically shown rising relative wages, along with rising supply, of more educated workers–but these trends are understood to partially reflect skill-biased changes in demands.) Therefore, these estimates identify a relatively short-run elasticity. A longer-run elasticity would presumably be larger. In particular, it will reflect technology's incentive to innovate towards the expanding groups (e.g., Caselli and Coleman (2006), Acemoglu (2007)). For questions of levels accounting, growth accounting, or longer-term impacts of policies on inequality, it is that longer-run elasticity that is relevant. Having said that, Ciccone and Peri (2005) estimate an elasticity of about 1.5 across schooling groups in the U.S. based on supply differences instrumented with state laws regulating child labor and compulsory schooling. They interpret this as a long-run elasticity. Also, Malmberg (2018) infers a comparably low elasticity based on export shares across countries in a gravity model

[1] This reasoning is incorporated more explicitly into general-equilibrium by Heckman, Lochner, and Taber (1998) and Johnson and Keane (2013), among others.

[2] By consensus, we mean setting aside the common practice of treating the elasticity as infinite, with schooling or skill groups perfect, but not equal, substitutes.

[3] Acemoglu and Autor (2011) extend the Katz and Murphy (1992) exercise through 2008 U.S. CPS data. They estimate elasticities of 1.6 to 2.9 depending on how flexibly they specify trends.

of trade driven by factor endowments, which we interpret as a fairly long-run elasticity.

We consider several avenues to gauge the elasticity of labor demand across schooling levels, especially with respect to long-run labor demand. Key to each avenue are the observed Mincer returns to schooling and how these returns vary across countries and over time. We consistently find evidence for an elasticity of demand across schooling groups of 4.0 to 5.0, much larger than the values typically employed in the literature.

In Section 2 we first review how an economy's Mincerian return reflects its quality of schooling, scarcity of more-educated workers, and the skill-bias of its technology. We show that relative scarcity of more-educated workers relates, not just to mean schooling, but also to the skewness of that distribution. We additionally allow for the skill bias in technology to reflect relative worker supplies, as in Caselli and Coleman (2006), Acemoglu (2007), or Hendricks and Schoellman (2019), to distinguish the short and long-run elasticities of labor demand across schooling types.

In Section 3 we relate Mincerian returns across countries to measures of both school quality and relative scarcity of more-schooled workers. Most previous work has focused on two groups of schooling. But cross-country data report fuller distributions of attainment, allowing us to empirically distinguish relative scarcity of more-schooled workers from the mean level of schooling. We employ three proxies for a country's school quality: One based on immigrants' earnings within the U.S. as in Schoellman (2012), one based on harmonized test scores, and one based on mean schooling attainment in a country. A low elasticity of substitution, such as employed in the literature, suggests an implausibly large cross-country relationship between development and quality of schooling, whereas an elasticity on the order of 4 aligns with measures of quality across countries. We also employ the cross-country measures of schooling scarcity and schooling quality to estimate the elasticity of substitution across schooling groups. This yields fairly high estimates of the elasticity, on the order of 4 or higher.

In Section 4 we employ growth accounting to examine the implications of the dramatic world-wide decline in relative scarcity of more-educated workers. This decline in relative scarcity has been associated with no drop in Mincerian returns. These joint observations imply a combination of (i) an elasticity of substitution that is quite high, much higher than 1.5, or (ii) extremely rapid technological change for workers with more schooling. But we show that overall productivity growth limits the role of the latter, pointing to an elasticity of substitution of at least 4.

Turning to the balance of the paper, in Section 5 we consider robustness in a couple dimensions, specifically with respect to the grouping of schooling types. Section (6) concludes, including considering the implications of employing an elasticity of demand across schooling groups on the order of 4, rather than 1.5, for purposes of income accounting across countries.

2 Skill, Scarcity, and Technology in Relative Wages

The basis of our analysis is the relationship between the return to schooling, measured in terms of wage premiums, and the underlying production structure that shape worker productivity. Throughout our analysis, we assume that labor markets are competitive, which links wages to productivity, and that workers of different schooling levels are imperfect substitutes, which links productivity to relative labor supply by schooling groups. Below, we study an environment where technology of production is a given and show how the wage premium reflects the scarcity of schooling in an economy. We discuss how the relevant measure of scarcity combines both the average schooling attainment and the skewness of the distribution of schooling. Then in Section 2.2 we allow the technology to expand in response to labor supply by schooling groups and derive the long-run relationship between scarcity and wage premium.

2.1 Scarcity and Wage Premium in the Short-run

Output of the economy is given by the Cobb-Douglas production function that takes physical capital, K, and human capital, H, as inputs.

$$Y = K^{\alpha} H^{1-\alpha}, \tag{1}$$

H aggregates labor supplied by workers with different schooling levels. Let S be the set of all schooling groups, and $L(s)$ be the number of workers with s years of schooling for all $s \in S$. The effective human capital is defined by the following constant elasticity of substitution (CES) aggregator.

$$H = \Big[\sum_{s \in S} \Big(e(s) L(s) \Big)^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}} \Big]^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}}, \tag{2}$$

where $\tilde{\varepsilon} > 1$ is the elasticity of substitution between schooling groups. $e(s)$ denotes the efficiency of schooling group s in production and it reflects the extent of their skills as well as the technical efficiency with which those skills are utilized:

$$e(s) = A(s) q(s). \tag{3}$$

$A(s)$ is the skill-specific technology level and $q(s)$ is the quantity of human capital associated with schooling level s. It reflects both the years and the quality of schooling.

Assuming that labor markets are competitive, the wage rate for a worker that belongs to schooling group s is:

$$w(s) = \frac{\partial Y}{\partial H} H^{\frac{1}{\tilde{\varepsilon}}} e(s)^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}} L(s)^{\frac{-1}{\tilde{\varepsilon}}}. \tag{4}$$

The wage rate is given by a combination of three components: the marginal product of aggregate human capital, which is constant for all schooling groups, the efficiency of schooling group in production, $e(s)$, and the size of the schooling group, $L(s)$. More efficient groups and groups that are scarce earn higher wages in the market.

The wage premium for schooling depends on the relative values of these components across schooling levels. To formalize this, define the Mincer return to schooling, m, as the coefficient of projection of log wage on years of schooling. Letting ϕ_a and ϕ_q denote the regression coefficients obtained by projecting $\ln A(s)$ and $\ln q(s)$ on years of schooling, the Mincer return is as follows:

$$m = \left(1 - \frac{1}{\tilde{\varepsilon}}\right)\left(\phi_a + \phi_q\right) + \frac{1}{\tilde{\varepsilon}} x. \tag{5}$$

$x = -Cov\big(\ln L(s), s\big)/Var(s)$ gives the average percentage decline in labor supply per year of schooling. It is a measure of the relative scarcity of skilled workers. A positive and large x indicates that, on average, supply of workers decreases rapidly with years of schooling.

In general, the scarcity of educated workers x will be negatively associated with average years of schooling, $\bar{s}$. For instance, if schooling attainment takes only one of two values from a fixed support $S = \{s_1, s_2\}$, then: $x = \ln\left(\frac{s_2 - \bar{s}}{\bar{s} - s_1}\right)/(s_2 - s_1)$. So $\bar{s}$ is a sufficient statistic for x. More generally, however, x will vary for given $\bar{s}$. We show in Appendix B that $x = \kappa \cdot \bar{s}$, where:

$$\begin{aligned} \kappa &= \frac{-1}{\sigma_s^2}\left[\sum_S \Big(\ln l(s)\Big) l(s) \left(\frac{s - \bar{s}}{\bar{s}}\right)\right] \\ &= \frac{\sum_S \left(l(s)|\frac{s-\bar{s}}{\bar{s}}|\right)}{2\sigma_s^2}\left[\mathbb{E}_{l(s)|\frac{s-\bar{s}}{\bar{s}}|}\Big(\ln l(s)\Big| s < \bar{s}\Big) - \mathbb{E}_{l(s)|\frac{s-\bar{s}}{\bar{s}}|}\Big(\ln l(s)\Big| s \geq \bar{s}\Big)\right]. \end{aligned} \tag{6}$$

κ measures the asymmetry of the schooling distribution $l(s) = L(s)/\sum_S L(s)$, hence it can be seen as a measure of skewness. To see this, note that the term in brackets captures, on average, by what percentage the density of the left tail exceeds that on the right. When l is symmetric, both $\ln l$ and $l(s)|\frac{s-\bar{s}}{\bar{s}}|$ are symmetric around $\bar{s}$, so $\kappa = 0$. More generally, we have $\kappa > 0$ if the distribution is right-skewed, and $\kappa < 0$ if it is left-skewed.

Equation (5) decomposes the Mincer return into a weighted average of two terms. The first term represents the relative efficiency of skilled groups in production. This *efficiency premium* combines the skill-bias in production technologies, ϕ_a, with the quality of schooling, ϕ_q. The second term represents the *scarcity premium*. The imperfect substitution of skill groups leads to a downward sloping demand curve for each schooling group, resulting in higher wages for skill groups that are short in supply. The weights on the two terms are determined by the inverse elasticity of skill substitution. The less substitutable skills are (lower $\tilde{\varepsilon}$), the higher is the scarcity premium and the lower is the efficiency premium. Thus, at lower values for $\tilde{\varepsilon}$, the Mincer return primarily reflects a group's scarcity. At the extreme case when $\tilde{\varepsilon} = 1$, for instance, efficiency gains from schooling are fully offset by the decline in marginal product along the demand curve. As a result, earning shares of different schooling groups are constant and the wage premium is determined solely by the relative scarcity of workers in each group. If, instead, skills are easily substitutable then the Mincer return largely reflects efficiency gains from schooling–with aggregate human capital linear in efficiency units, and scarcity irrelevant, as $\tilde{\varepsilon} \to +\infty$.

This suggests that the elasticity of skill substitution can be inferred from the relation between the Mincer return and skill scarcity in the data. There are two empirical challenges. First, the scarcity of skills may be related to the quality of schooling in a country. If skills are scarce in countries with poor education systems, then the Mincer return may be seemingly unresponsive to scarcity of skills, even when the underlying elasticity of skill substitution may be low. We tackle this empirical challenge in Section 3 by approximating school quality with available measures in the data.

The second empirical challenge is that the efficiency gains through accessing technology from schooling may be systematically correlated with the scarcity of schooling. In a model of endogenous technological investments, Caselli and Coleman (2006) show that countries that are abundant in skills choose to specialize in technologies that favor skilled workers. Thus, more workers of higher schooling in a country may fail to reduce the country's Mincer return, not because scarcity does not matter (as would be the case when skills are highly substitutable), but because technological investments offset the relative supply effects on marginal products. This concern is particularly relevant for long-run economic growth, which is driven not only by accumulation of skills but also by technological choices. It is equally relevant for cross-country comparisons of the observed Mincer return, which is likely a culmination of past investments in both skills and technology. For these reasons, we next present a model where technological choices are endogenous to skill endowments and develop the *long-run* relationship between the wage premium for skills and the scarcity of skills.[4]

2.2 Technology and the Long-run Elasticity of Substitution

We follow Caselli and Coleman (2006) in our formulation of skill-specific technology investment. Technologies are chosen by firms among a set of possibilities defined by a technology

[4] Hendricks and Schoellman (2019) is a recent example of applying the long-run substitutability among schooling groups in development (levels) accounting.

frontier, given the skill endowments in the country.

Formally, given the wage rates $\{w(s)\}_s$ and the interest rate R, firms rent capital K, adopt technologies $A(s)$, and hire workers $L(s)$ from each skill group to maximize profits $(K^{\alpha}H^{1-\alpha} - \sum w(s)L(s) - RK)$ subject to the following technology frontier.

$$\sum_{s\in S}\Big[\gamma(s)A(s)\Big]^{\omega} \leq B. \tag{7}$$

The parameters $\omega > 0$ and $\{\gamma(s)\}_s$ determine the technical trade-off between technologies associated with different schooling groups. If $\gamma(s)$ is increasing in s, then investment in technologies assigned to skilled workers is technically more costly. ω is the elasticity of technical transformation between $A(s)$, and it shows how fast the marginal cost of technological enhancement rises with the level of the technology. Following Caselli and Ciccone (2013), we assume here that ω and $\gamma(s)$ are common to all countries. But we relax this assumption for some of our empirical work below. B_c determines the level of the technology frontier, which may differ by country.

To focus on the equilibrium where firms choose $A(s) > 0$ and hence $L(s) > 0$ for all $s \in S$, we make the following assumption on the parameters.

$$\omega - \tilde{\varepsilon} + 1 > 0. \tag{8}$$

As shown in Appendix C.1, this assumption guarantees a symmetric equilibrium with interior input choices, so we can characterize the equilibrium using the optimality conditions of a representative firm. The optimal level of technology for schooling level s is described by:

$$A(s) = \left[\frac{q(s)L(s)}{\gamma(s)^{\frac{\omega}{\sigma}}}\right]^{\frac{\sigma}{\omega-\sigma}} Q^{\frac{1}{\omega-\sigma}}, \tag{9}$$

where $\sigma = (\tilde{\varepsilon}-1)/\tilde{\varepsilon}$, which is positive when $\tilde{\varepsilon} > 1$, as we assume here. The technology choices generally depend on a scale effect, $Q = B/H^{\sigma}$, determined by the country's aggregate human capital stock and the level of their technological capacity, B. The relative technology across skill groups depend on the endowments of human capital for each group, $q(s)L(s)$, relative to the marginal cost of technology for that skill group, $\gamma(s)$. If a country is skill abundant, it is optimal to invest more heavily in skill-biased technology. The extent of the bias depends on the slope of $\gamma(s)$ with respect to s. If $\gamma(s)$ increase more with s, a larger quantity of skilled workers are needed before skill-biased technologies become optimal.

As shown in Hendricks and Schoellman (2019), equation (9) implies that the equilibrium allocation and prices in the labor market are equivalent to those given by the optimality conditions of a representative firm with the following alternative aggregator.[5]

$$H = B^{\frac{1}{\omega}} \left[\sum_{s \in S} \left(\frac{q(s)}{\gamma(s)} L(s) \right)^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{10}$$

where the elasticity of substitution is

$$\varepsilon = \frac{\omega\tilde{\varepsilon} - \tilde{\varepsilon} + 1}{\omega - \tilde{\varepsilon} + 1}. \tag{11}$$

The assumption in (8) guarantees that ε is finite and $\varepsilon > \tilde{\varepsilon}$, for all $\tilde{\varepsilon} > 0$. Since we assume $\tilde{\varepsilon} > 1$, we also have $\varepsilon > 1$. This elasticity of substitution can be considered to be the long-run elasticity of substitution where technology is endogenous. The wage rate for a worker with s years of schooling implied by (10) is

$$w(s) = \frac{\partial Y}{\partial H} H^{\frac{1}{\varepsilon}} B^{\frac{\varepsilon-1}{\varepsilon\omega}} \times \left(\frac{q(s)}{\gamma(s)} \right)^{\frac{\varepsilon-1}{\varepsilon}} L(s)^{-\frac{1}{\varepsilon}}, \tag{12}$$

and the associated log-wage premium for education in country c is:

$$m = \frac{\varepsilon - 1}{\varepsilon} \left(\phi_q - \phi_\gamma \right) + \frac{1}{\varepsilon} x. \tag{13}$$

ϕ_γ is the log-projection of the world's skill-specific technology costs, $\gamma(s)$, on years of schooling. ϕ_q is the log-projection of quality $q(s)$ on years of schooling in each country. Because countries face the same marginal trade-offs in technology investment across schooling groups, ϕ_γ is common to all countries. It is positive if technology is relatively more costly for higher skill groups: $\gamma'(s) > 0$.

Equation (13) forms the basis of our cross-country analysis presented in Section 3. It shows that the differences in the wage premium for education across countries is a weighted average of the skill premium, which reflects differences in the quality of schooling net of skill-bias in technology costs, $\phi_q - \phi_\gamma$, and skill scarcity, x, as in the short-run version of the wage premium in equation (5). The main difference between the two equations is that the wage premium for scarcity is lower in the long-run since $\varepsilon > \tilde{\varepsilon}$. This is because the negative wage effect of an increase in the supply of a skill group is partly offset by improvements brought to the technology of production assigned to that skill group. A comparison of Mincer returns across countries with the scarcity of skills is therefore informative of the long-run elasticity of skill substitution. Lower values of ε represent harder skill substitution in the long run and results in a higher valuation of scarcity. Consequently, the cross-country differences in the wage premium, indicated by the Mincer return, should be driven primarily by scarcity of skills.

[5] A derivation is provided in Appendix C.2.

3 Cross-Country Estimates of the Elasticity of Substitution

We first investigate the patterns in skill scarcity and wage premium across countries. This suggests an approach for estimating the elasticity of substitution, ε, which we conduct employing alternative measures or controls for quality of schooling.

3.1 Patterns in Scarcity and Mincer Returns across Countries

To investigate the cross-country patterns in skill scarcity and wage premium, we estimate skill scarcity using data on educational attainment by country provided in Barro and Lee (2013). The data are a panel of 153 countries covering the period 1950 to 2010 at 5 yearly intervals. It reports population frequencies over 7 educational categories by broad age groups.[6] We restrict the sample to those ages 25 to 54 to capture a working age population. We assign years of schooling to each attainment level using data from UNESCO on the duration of educational categories in each country.

To measure skill scarcity, for each country in each year, we regress the (log) size of the population in each schooling category on the years of schooling for that category. The negative of the estimated coefficient, which shows the average percentage-point decline in labor supply per year of schooling, is our measure of skill scarcity, x, in each country for each year. Our benchmark case divides workers into four groups: i) completed primary or less, ii) some secondary schooling, iii) completed secondary, and iv) any tertiary (e.g. college) education.[7] We show robustness of our results for alternative groupings in Section 5.1, dividing countries into two, three, or more than four groupings.

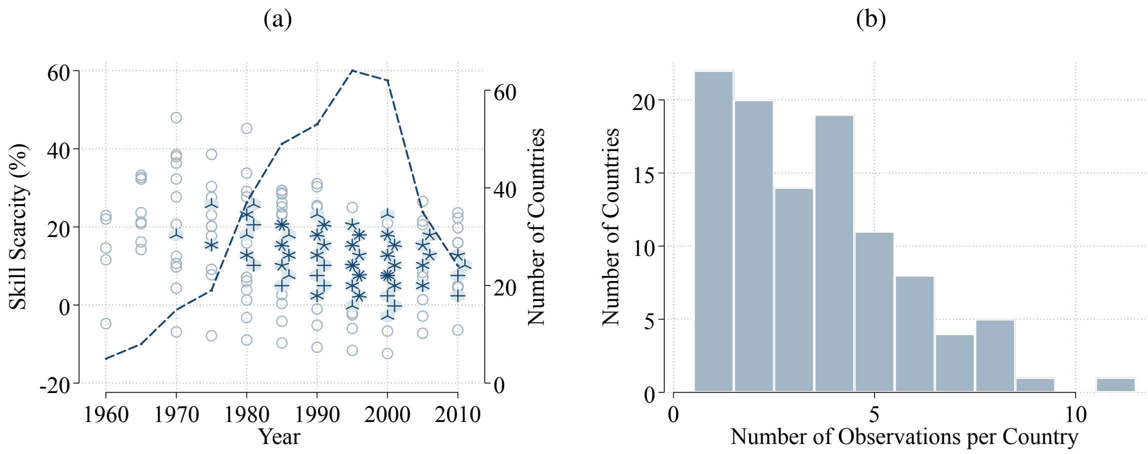
The data on Mincer returns are obtained from Psacharopoulos and Patrinos (2018), which is a meta-analysis of Mincer regressions covering many countries for various years. We merge the two datasets to obtain an unbalanced panel sample of 105 countries for the years 1960 to 2010 containing 371 observations with data on both x and the Mincer return. For more details on the data sets we employ, please see the data appendix.

Figure 1 shows the sampling distribution of scarcity in our merged sample. Panel (a) shows the sunflower graph of the distribution of skill scarcity in our sample. Each circle or flower petal indicates an observation. The observations are mainly concentrated between the years 1975 to 2010. Skill scarcity trends downward due to improvements in educational attainment around the world. Panel (b) shows the frequency distributions of observations per country. 14 countries have single year observations. The remaining countries have at least two observations

[6] In Section 5.2, we show that using employment shares instead of population shares of schooling groups does not affect our findings.

[7] For our 105 country sample below, the average shares by group are respectively 47%, 20%, 22%, and 11%.

Figure 1: Skill Scarcity across Countries.



Notes.– Panel (a) shows the sunflower graph of the distribution of skill scarcity in our sample. Each circle or petal indicates an observation. The dashed line shows the total number of observations for each year (right axis). Panel (b) shows the number of observations per country in the sample.

during the sample period.

Panel (a) in Figure 2 shows the distribution of scarcity against average labor productivity, as measured by output per worker. Perhaps not surprisingly, countries that are abundant in skilled workers (low x) have higher levels of average labor productivity. If wage premiums are driven by scarcity of skills, then the abundance of skills should translate to lower wage premiums in these countries. Panel (b), which plots the distribution of Mincer return estimates from Psacharopoulos and Patrinos (2018) by labor productivity, shows instead that the wage premium is remarkably flat. Absent differences in ϕ_q or ϕ_γ, this would suggest that the elasticity of skill substitution cannot be very low.

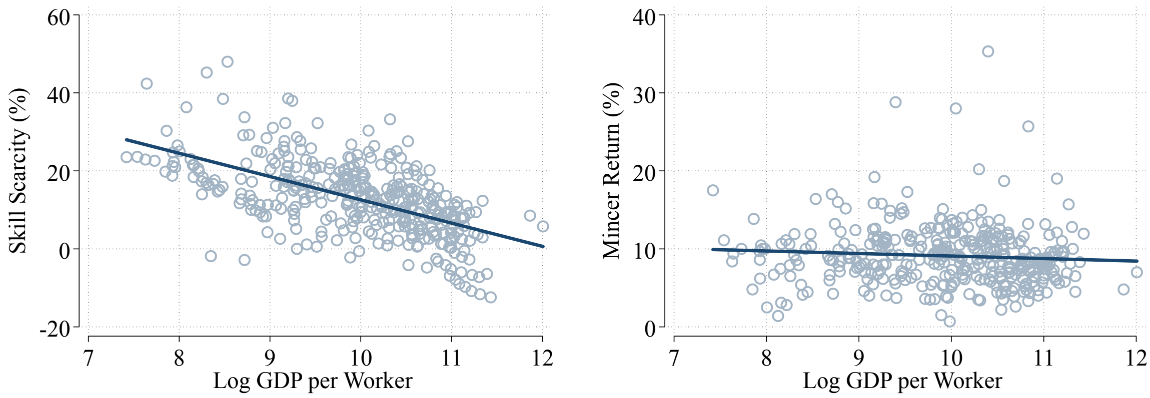
Based on equation (13), one way to reconcile the patterns observed in Figure 2 with a low elasticity of skill substitution is to have a sufficiently strong negative relationship between school quality, ϕ_q, and skill scarcity, x. In other words, the quality of education has to be sufficiently lower in countries where higher-schooled workers are scarce. In that case, the differences in the skill premium and in the scarcity premium could offset each other, resulting in a flat wage premium across countries. To see this, rearrange equation (13) to back out the school quality:

$$\tilde{\phi}_{q,c}(\varepsilon) = \frac{\varepsilon}{\varepsilon - 1}\left[m_c - \frac{1}{\varepsilon}x_c\right] + \phi_\gamma. \tag{14}$$

If the Mincer return is to remain stable as skill scarcity varies, as we see in Figure 2, countries where skills are scarce must offer steeper skill gains from schooling, especially when the degree of skill substitution, ε, is low.

Figure 2: Mincer Return and Skill Scarcity across Countries.

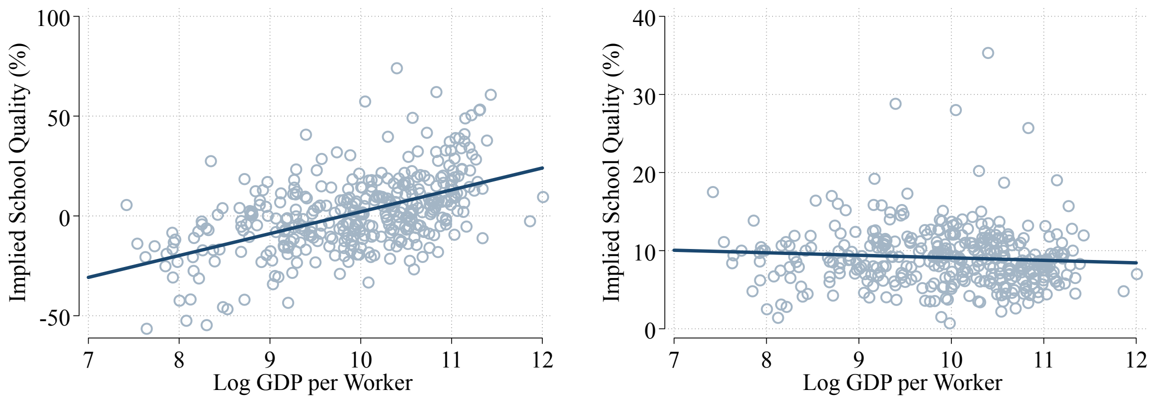
(a) Skill Scarcity (%) (b) Mincer Return (%)



Notes.– Data on the Mincer return ($\widehat{m}$) is taken from Psacharopoulos and Patrinos (2018). Skill scarcity (x) is obtained from authors' calculations based on Barro and Lee (2013). Log real GDP per worker (y) is obtained from Penn World Tables 9.1.

Figure 3: Implied School Quality $\tilde{\phi}_{q,c}(\varepsilon)$.

(a) $\varepsilon = 1.5$ (b) $\varepsilon = \infty$



Notes.– The dots show the school quality ($\tilde{\phi}_q$) in each country that is consistent with the Mincer return ($\widehat{m}$) and skill scarcity (x) for $\varepsilon - 1.5$ (Panel a), and $epsilon = infty$ (Panel b). The blue, solid lines depict the projection of these school qualities on GDP per worker. Data on the Mincer return ($\widehat{m}$) is taken from Psacharopoulos and Patrinos (2018). Skill scarcity (x) is obtained from authors' calculations based on Barro and Lee (2013). Log real GDP per worker (y) is obtained from Penn World Tables 9.1.

To explore this line of argument, we compute the schooling quality that is consistent with the observed Mincer return and skill scarcity for each country. We follow Caselli and Coleman (2006) and Hendricks and Schoellman (2019) in assuming common parameters $\gamma(s)$ across countries.[8] Therefore, without loss of generality, we set the common term $\phi_\gamma = 0$. The results for $\varepsilon = 1.5$ are plotted in Panel (a) of Figure 3. The dots are the computed values of $\tilde{\phi}_{q,c}(\varepsilon)$ for each country and year by its GDP per worker. The regression line is shown in solid blue.

The differences in school quality needed to explain the cross-country patterns in scarcity and wage premium are immense. On average, one percent increase in labor productivity has to be associated with 11 basis point increase in the quality of schooling *per year*. In 2000, for instance, the interquartile range of labor productivity across countries was 1.7 log points. This implies that a high school graduate, with 12 years of schooling, in the richer country has 9 times the human capital of a similar worker in the poorer country and a college graduate with 16 years of schooling has 20 times the human capital. Panel (b), on the other hand, plots the results for $\varepsilon = \infty$. According to equation (14), the implied school quality equals Mincer return when schooling groups are perfect substitutes. The result implies that richer countries, on average, have lower quality of schooling, which is also at odds with measures of school quality as we find below. These two examples point to our main strategy to estimate the elasticity of substitution among schooling groups—measuring the variation of school quality across countries.

3.2 Measuring Quality from Wages of U.S. Immigrants

To set a point of reference, we pull data from Schoellman (2012) who provides estimates of school quality based on immigrants in the US. The estimates reflect the wage premium for education in the US among immigrants originating from the same country. Because these immigrants had obtained their schooling prior to immigration and because their wages are set in the same labor market, the variation in wage premiums across countries of origin reflects the relative quality of schooling there rather than market-specific factors such as technology or skill scarcity. Specifically, letting c denote the country of origin, we assume that the aggregate human capital in US is given by:

$$H_{US} = \left\{ \sum_{s \in S} \left[\sum_{c \in C} e_c(s) L_c(s) \right]^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}} \right\}^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}}.$$

This formulation assumes that immigrants from different countries are perfect but potentially unequal substitutes. The implied US wage premium is,

$$m_c^{US} = \zeta + \phi_{q,c},$$

[8] That is, this exercise allows differences in technologies attached to workers in the two countries through the directed technology described in 2.2, but not differential values for ϕ_γ.

where ζ is a US-specific constant reflecting the skill-bias of technology and skill scarcity in the US. (See Appendix D for the derivations.)

We regress the school quality levels estimated by Schoellman (2012) for each country on output per worker and average school attainment respectively. The results are displayed with red dashed lines in Figure 4, with the scatter plots of original quality measures in triangles. Solid blue lines show the income-quality gradient associated with the assumed elasticity of substitution. The estimates based on US immigrants show a much smaller rise in school quality by income (panels a, c, and e) and by educational attainment (panels b, d, and f). A one percent increase in income per worker is associated with a 1.4 percentage point higher school quality, about an order of magnitude smaller than the 11ppt suggested by the solid blue line in Panel a. Considering the interquartile range in 2000 once again, a high school graduate is 1.3 times the human capital in the richer country and a college graduate 1.5 times the human capital. While these are significant differences in quality, they are much smaller than the ratios that are necessary to justify an elasticity of skill substitution of 1.5. Similar pattern holds when projecting on educational attainments in Panel b.

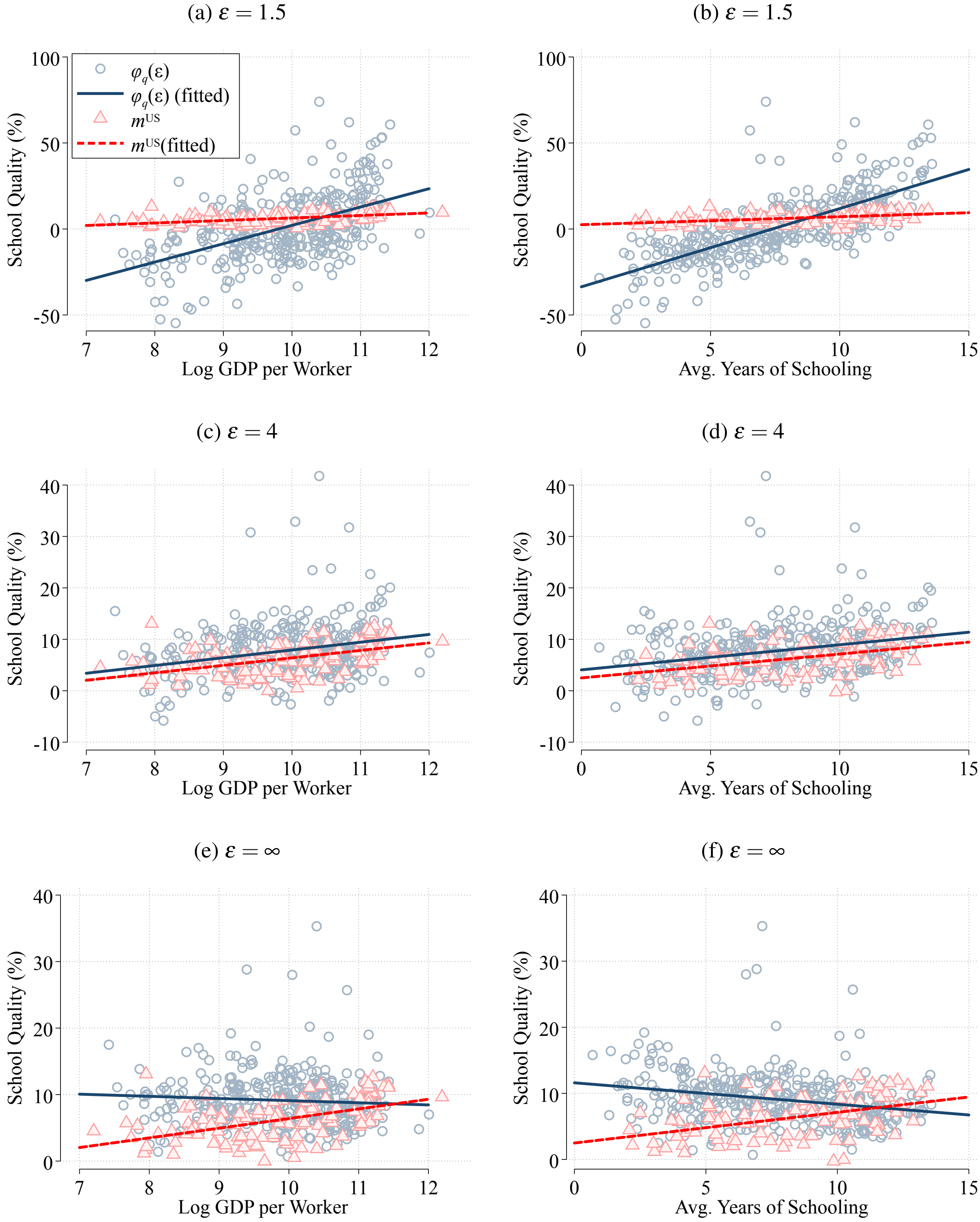
One might be concerned that selection into immigrating to the US might affect the estimates of the quality of schooling. Note that, for this to change the quality-income gradient shown by the slope of the red dashed line in Figure 4, a systematic, cross-country variation in the differential selection by schooling is necessary on unobserved traits of productivity. The true quality-income gradient would be steeper if, for instance, the US received the best college graduates from poor countries and the worst ones from rich countries. Hendricks and Schoellman (2018) examine the selection patterns in unobserved traits using panel data on immigrants to the US and do not find any correlation between selection and education within an income group. We expect, therefore, that selection of immigrants is unlikely to change our conclusions in Figure 4 in a significant way.[9]

The remaining panels repeat the exercise for higher values of ε. A higher elasticity of substitution generally requires smaller gaps in school quality across countries in order to explain the patterns in Figure 2. When skills are perfect substitutes ($\varepsilon = \infty$), scarcity has no effect on the wage premium and the Mincer return shows only school quality differences. Given the weak relationship between labor productivity and the Mincer return in Panel (e), the school quality is slightly lower in rich countries. This too is at odds with measures based on US immigrants, which suggest that countries with higher income per worker have schools of better quality on average. The two regression lines coincide when $\varepsilon = 4$ in Panel (c).

Figure 4 hints at an estimation approach that uses the measures of school quality. We can

[9]The limited cross-country coverage in the panel survey of new immigrants used by Hendricks and Schoellman (2018) prevents us from deploying it in our analysis.

Figure 4: School Quality and the Elasticity of Skill Substitution



Notes.– Figure shows the school quality ($\tilde{\phi}_q$) in each country that is consistent with the Mincer return ($\widehat{m}$) and skill scarcity (x) for different values of the elasticity of skill substitution (ε). The blue, solid line depicts the projection of these school qualities on GDP per worker. The red, dashed line shows a comparable projection using school quality estimates ($\widehat{m}^{US}$) reported by Schoellman (2012).

Table 1: Estimates of Substitutability (Quality measured from U.S. Immigrant Wages)

	(1)	(2)	(3)
Slope	0.25	0.22	0.20
	(0.05)	(0.05)	(0.07)
$\widehat{\varepsilon}$	4.05	4.45	4.94
Interval	[2.81, 7.27]	[2.99, 8.69]	[2.98, 14.43]
Estimator	OLS	GMM	GMM
Instruments	–	x_c	$\bar{s}_c$, y_c

Note.– Table shows results from least squares regressions of $\widehat{m}-\widehat{m}^{US}$ on $x-\widehat{m}^{US}$. Column (1) shows the OLS estimate; Columns (2) and (3) are estimated by GMM using instruments as stated. Data on the Mincer return ($\widehat{m}$) are from Psacharopoulos and Patrinos (2018). Skill scarcity (x) and average schooling attainment ($\bar{s}$) are obtained from authors' calculations based on Barro and Lee (2013). Log real GDP per worker (y) is obtained from Penn World Tables 9.1. Data on the Mincer return on US immigrants ($\widehat{m}^{US}$) is taken from Schoellman (2012). Robust standard errors are reported in parentheses. Sample size is 51 countries for all columns.

rearrange (13) to obtain:

$$m-\phi_q=\frac{1}{\varepsilon}(x-\phi_q)-\frac{\varepsilon-1}{\varepsilon}\phi_\gamma\,. \tag{15}$$

This provides a simple regression approach for estimating the magnitude of $1/\varepsilon$, conditional on having a measure for how quality of schooling, ϕ_q varies across countries. We employ two alternative measures for ϕ_q, one based on the data on U.S. immigrants' earnings from Schoellman (2012), $\widehat{\phi}_q=\widehat{m}^{US}$, the other based on comparisons of standardized test scores across countries (discussed in Section 3.3), $\widehat{\phi}_q=\widehat{\phi}^{PISA}$.

The error term in estimating (15) reflects measurement errors in the variable as well as heterogeneity in ϕ_γ. Either proxy for ϕ_q, based on U.S. immigrants' earnings or test scores, are clearly imperfect. This implies direct estimates of (15) may suffer from attenuation bias that biases an estimate of ε toward one. Therefore, we estimate (15) by instrumenting for $(x-\phi_q)$ as well as by OLS. Any bias from cross-country variations in ϕ_γ depend on its covariance with $(x-\phi_q)$, or those variables acting as instruments. In particular, if richer countries exhibit lower values for ϕ_γ, this will downwardly bias our estimate of $\frac{1}{\varepsilon}$ – so upwardly bias $\widehat{\varepsilon}$. That concern motivates our additional exercises in Section 3.4 and Section 4 as discussed below.

Table 1 shows the regression results using 51 countries from Schoellman's sample whose Mincer returns and schooling distributions are observed in our main sample in 2000. For each regression, we first show the estimated slope and its robust standard error. Then we take the inverse of the slope as a point estimate for the elasticity of substitution ($\widehat{\varepsilon}$). Below the point estimate, we show the inverse of the 95% confidence interval of the slope. Note that every hypothesis $H_0: 1/\varepsilon=1/\varepsilon_0$ with ε_0 outside of this interval will be rejected with 95% confidence level and *vice versa*, so the interval is the set of values not rejected.

The first column of Table 1 gives the OLS results. The estimated value of the elasticity is about 4, with a 95% confidence interval of 2.8 to 7.3. Given the U.S.-based Mincer returns from Schoellman (2012) will reflect measurement error, in Column two we instrument for the right-hand-side variable, $x - \widehat{m}^{US}$, with schooling scarcity in a country, x. Thus these estimates only reflect the U.S.-based Mincer returns to the extent they project on scarcity, x in that country. The Column 2 estimate for ε is higher at 4.5 (confidence interval 3.0 to 8.7), consistent with attenuation bias pushing the OLS estimate toward one. In Column 3 we instrument for the entirety of $x - \widehat{m}^{US}$ based on a countrys average schooling attainment and real GDP per worker. This yields an even higher estimate for ε of 4.9 (confidence interval 3.0 to 14.4). These results are consistent with our observation from Figure 4 – the commonly used elasticity of substitution, $\varepsilon = 1.5$ or ∞, are clearly rejected by our estimation. Indeed, our estimation shows even $\varepsilon = 2.5$ is too low to explain the cross-country patterns of Mincer returns, school quality, and skill scarcity.[10]

Conditional on using instrumental variables to estimate equation (15), it is not necessary that a country have data on all three variables: $\widehat{m}$, x, and $\widehat{m}^{US}$ in order to contribute in estimating $\widehat{\varepsilon}$. This is useful, as we can observe the variables x and $\widehat{m}$ for a much broader sample of countries than our measure $\widehat{\phi}_q$ based on immigrant earnings. For instrument z, an instrumental variables estimator of equation (15) can be written as:

$$\widehat{\varepsilon^{-1}} = \frac{b_{z,\, \widehat{m} - \widehat{\phi}_q}}{b_{z,\, x - \widehat{\phi}_q}} = \frac{b_{z,\, \widehat{m}} - b_{z,\, \widehat{\phi}_q}}{b_{z,\, x} - b_{z,\, \widehat{\phi}_q}} . \tag{16}$$

$b_{z,y}$ denotes the slope of variable y with respect to the instrument z. That is, $b_{z,y} = (\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{y}$ (with both y and z demeaned). The variables $\widehat{m}$ and $\widehat{\phi}_q$ denote, respectively, a country's domestically measured Mincer return and our measure of its school quality (based on U.S. immigrants or test scores as we turn to shortly). A country subscript c is implicit. The estimator is a combination of the slopes of the three variables $\widehat{m}$, $\widehat{\phi}_q$, and x with respect to z, where these need not be estimated on a common sample of countries. To estimate slope parameters for $\widehat{m}$ and x, we merge the Psacharopoulos and Patrinos (2018) data on the Mincer returns with the Barro and Lee (2013) data on educational attainment. To estimate the slope parameter for $\widehat{\phi}_q$, we merge the Schoellman (2012) data on school quality with the Barro and Lee (2013) data. For an instrument we consider a country's average schooling attainment or its real GDP per worker.

Table 2 shows the results for these auxiliary regressions, where Panel A uses average attainment of schooling ($\bar{s}_c$) as the instrument and Panel B uses log real income per person (y_c). The first column of each panel projects the Mincer return on the instrument. It shows that an extra year of education is associated with 0.32ppt drop in the wage premium, and an extra

[10]Weighting countries by their relative workforces as of 2000 yields somewhat higher estimates for the elasticity, with $\widehat{\varepsilon}$ ranging from 6 to 8 for the regressions in Table 1.

Table 2: Regression Results for Multiple-Sample Estimation

(a) $\widehat{\varepsilon} = 3.96$

	(1) $\widehat{m}_c$	(2) x_c	(3) $\widehat{m}_c^{US}$
$\bar{s}_c$	–0.32	–2.65	0.46
	(0.06)	(0.09)	(0.09)
R^2	0.09	0.81	0.20
N	371	371	101

(b) $\widehat{\varepsilon} = 4.13$

	(1) $\widehat{m}_c$	(2) x_c	(3) $\widehat{m}_c^{US}$
y_c	–0.27	–5.64	1.45
	(0.21)	(0.43)	(0.24)
R^2	0.03	0.46	0.25
N	367	367	116

Note.– Table shows results from least squares regressions. Data on the Mincer return estimates ($\widehat{m}$) is taken from Psacharopoulos and Patrinos (2018). Skill scarcity (x) is obtained from authors' calculations based on Barro and Lee (2013). Log real GDP per worker (y) is obtained from Penn World Tables 9.1. Data on the Mincer return on US immigrants ($\widehat{m}^{US}$) is taken from Schoellman (2012). For both panels, Columns (1) and (3) include controls for year indicators. Robust standard errors are reported in parentheses.

percentage point of income is associated with 0.27ppt drop. The second column projects the skill scarcity on years of schooling and income. Across countries, an additional year of average schooling is associated with a 2.65ppt decline in skill scarcity, and an additional percentage point of real income is associated with 5.64ppt decline.[11] The third columns shows that, on average, countries with higher educational attainment and real income have higher measured school qualities ($\widehat{m}^{US}$), the gradient of efficiency unit with respect to schooling. An additional year of schooling is associated with 0.46ppt increase in school quality, and one percent higher real income predicts 1.45ppt increase.

These findings suggest a non-trivial role for skill scarcity in explaining the wage premiums for schooling. The results in Columns (1) and (3) seem to be in contradiction as they suggest that countries with lower education attainments and/or lower incomes have higher wage premiums for schooling despite their lower schooling quality. The second column proposes an explanation. It indicates that more educated workers are more scarce in countries where the average years of schooling is low and where real income is low. The scarcity of skills can potentially lead to higher wage premiums in poor countries if they more than offset the lower efficiency gains from schooling (i.e. school qualities). This in turn implies that labor among schooling groups must be imperfect substitutes. Combining these estimates in equation (16), once again, yields an elasticity of substitution around 4.

3.3 Measuring Quality from International Student Assessments

Supplemental evidence on school quality from the Programme for International Student Assessment (PISA) reinforce these findings. PISA's objective is to evaluate educational systems

[11]The R^2 shows average attainment and scarcity are strongly correlated, with correlation coefficient of 0.8.

around the world by testing the scholastic abilities of 15-year old students along three dimensions: mathematics, science and reading. We construct two school quality measures based on the micro-level data on test results provided by the OECD for the 2015 wave of tests. The benchmark results reported here use the test scores for mathematics.[12]

PISA specifically targets 15-year old students in each country. Students will be at different grades when they take the test, however, if the school starting age is dependent on the month of the year they were born, or if there are other differences in schooling systems across regions, or, even schools. Our first measure of quality uses the variation in schooling grades at the time the student takes the test. In each country, we regress the test score on the grade year in which the test was taken, controlling for gender. We restrict the sample to native-born students who never repeated a grade. The coefficient on the grade year gives the return to a year of schooling in terms of the test score and forms the basis of our quality measure. Comparing the resulting measure, which is in units of standardized test scores, with the school quality, $\tilde{\phi}_{q,c}(\varepsilon)$, which is in wage units, requires a market value for the test score. We calibrate this value using US data on the wage return to standardized test scores. To that end, we first divide the marginal test score attributable to a year of schooling by the standard deviation of the test score in US. This allows us to express schooling quality in terms of a unit standard deviation of US test scores. Then, using data from the 1979 cohort of the National Longitudinal Survey of Youth, which contains results on Armed Forces Qualification Tests, we find that an increase in the test score by one standard deviation is associated with a 15 percent increase in wages.[13]

The second measure of school quality is the ratio of the test score to the modal grade year in each country when the test was taken. Similar to our first measure, we divide the resulting per-school-year score by the standard deviation of US test scores and valorize it at 15 percent.[14]

Panels (a) and (c) in Figure 5 show the scatter plot of our first test-based quality measure (in triangles) against GDP per worker across 57 countries where data on scarcity, PISA tests and the Mincer return are all available.[15] The red dashed line shows the regression line. The blue dashed line shows the regression line corresponding to our second test-based quality measure. Both lines indicate that school quality is positively correlated with labor productivity of a country. The dark solid line shows the fitted school quality levels that are consistent with an elasticity of substitution of 1.5 in panel (a) and 4 in panel (c), as in Panel (a) and (c) of Figure

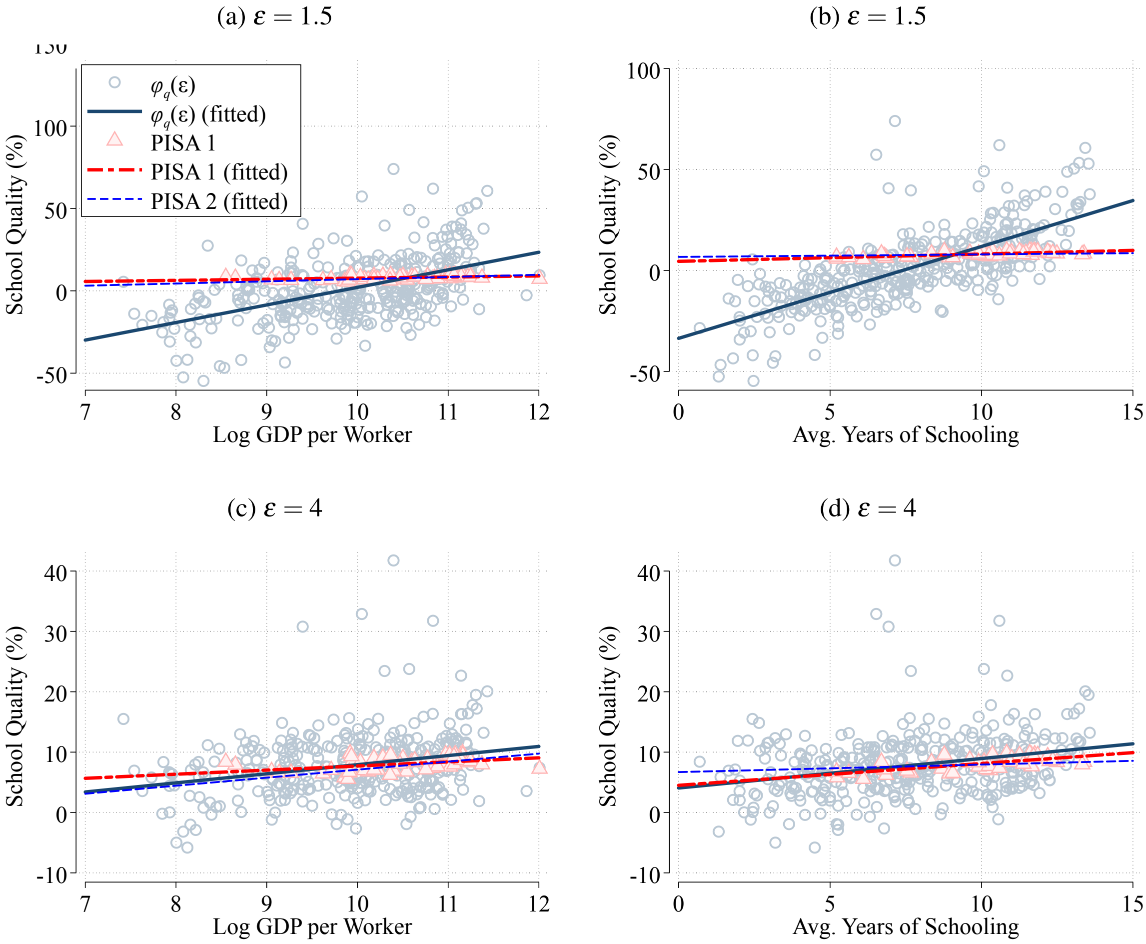
[12]We pick mathematics because we think the content is more comparable across countries, and because it correlates more strongly with average educational attainment across countries. Results based on other fields yield higher estimates of the elasticity of substitution. Our choice is therefore conservative.

[13]The estimate is based on the NLSY's nationally representative sample, but restricted to men with at least 15 years of work experience. The wage return is slightly lower among younger workers due to employer uncertainty regarding worker ability (Altonji and Pierret, 2001).

[14]The implicit assumption behind this measure is that the test score prior to schooling is zero. If developed countries have better pre-school training, then our measure is biased up for these countries, and so is the implied elasticity of substitution.

[15]The GDP per worker is the average value of all available years in our sample between 1995 and 2010.

Figure 5: School Quality from PISA



Notes.– PISA 1 and PISA 2 are measures of school quality based on test results from the Programme for International Student Assessment and author's calculations. See text for details.

Table 3: Estimates of Substitutability (PISA-based Measure of School Quality)

	(1)	(2)	(3)
Slope	0.24	0.21	0.24
	(0.09)	(0.09)	(0.10)
$\widehat{\varepsilon}$	4.25	4.83	4.15
Interval	[2.39, 19.34]	[2.59, 34.98]	[2.31, 20.85]
Estimator	OLS	GMM	GMM
IV	–	x_c	$\bar{s}_c$, y_c

Note.– Table shows results from least squares regressions of $\widehat{m} - \widehat{\phi}^{PISA}$ on $x - \widehat{\phi}^{PISA}$. Column (1) shows the OLS estimate, and Column (2)-(4) are estimated by general method of moments (GMM) with different instrumental variables. Data on the Mincer return ($\widehat{m}$) is taken from Psacharopoulos and Patrinos (2018). Skill scarcity (x) and average schooling attainment ($\bar{s}$) are obtained from authors' calculations based on Barro and Lee (2013). Log real GDP per worker (y) is obtained from Penn World Tables 9.1. See text for an explanation of school quality estimates ($\widehat{\phi}^{PISA}$) based on the PISA test scores. Robust standard errors are reported in parentheses. Sample size is 51 countries for all columns.

2. Recall from Figure 2 that steeper quality-schooling gradients indicate lower degrees of substitution. That the fitted line of quality measures is parallel to the blue solid line when $\varepsilon = 4$ therefore implies that the elasticity of substitution implied by the second test-based measure is close to 4. Indeed, using equation (16), the implied elasticity estimate is 4.36. The flatter, red dashed line implies an elasticity of substitution of 6.66 using the first test measure.

Panels (b) and (d) show the scatter plot of school quality and average years of schooling in a country. Countries with higher measured quality have higher educational attainments on average. The schooling gradient of quality measures can be used to estimate the elasticity of substitution using equation 16. Steeper quality-schooling gradients indicate lower degrees of substitution. The resulting elasticity estimates are 4.42 when using the first PISA quality measure (red-dash line) and 6.24 when using the second (blue dash line).

Table 3 shows the estimates of substitutability obtained by regressing $\widehat{m} - \widehat{\phi}^{PISA}$ on $x - \widehat{\phi}^{PISA}$. The results correspond to our preferred measure of school quality based on grade variation among test takers. As in Table 1, the first column is estimated by OLS and the others are estimated by GMM using different instruments to correct for the attenuation bias. The estimated elasticity of substitution varies between 4.15 and 4.8 depending on the specification. The corresponding lower bound of the 95% confidence interval runs from 2.3 to 2.6. These results are consistent with those obtained in Table 1 using school quality measures based on US immigrants. It is reassuring that the two very different approaches to measure school quality yield commensurate degrees of labor substitutability between schooling groups.

Table 4: Regression Results

	(1)	(2)	(3)	(4)
Dependent variable:	$\widehat{m}_c$	$\widehat{m}_c$	$\widehat{m}_c$	$\widehat{m}_c$
$\bar{s}_c$		0.25		0.58
		(0.13)		(0.63)
x_c	0.14^{***}	0.21^{***}	0.18	0.28^{*}
	(0.02)	(0.05)	(0.10)	(0.12)
Specification	levels	levels	trends	trends
N	371	371	83	83
$\widehat{\varepsilon}$	7.1	4.7	5.3	3.7
Interval	[5.6, 10]	[3.4, 7.7]	[2.6, 11.1]	[2.0, 20]

*** p<0.01, ** p<0.05, * p<0.1

Note.– Table shows the regression of the Mincer return to schooling on average years of schooling, $\bar{s}$, and scarcity of skills, x. The first two columns control for year indicators. Robust standard errors are reported in parentheses.

3.4 Accounting for Quality with Average Attainment

The findings in Sections 3.2 and 3.3 show that average years of schooling in a country is positively correlated with the quality of schooling. This suggests that the average years of schooling can potentially serve as a proxy for the unobserved school quality in equation (13). More broadly, it reflects the differences in efficiency of schooled labor, potentially including those associated with ϕ_γ, as skill-biased reductions in technology costs would encourage investment in schooling in the long-run. As an alternative to our estimates based on quality proxies from immigrants' earnings or from test score, we use average years of schooling as a control for the quality of education. More exactly, we regress the Mincer return on skill scarcity controlling for the average years of schooling in the merged sample that combines data from Barro and Lee (2013) and Psacharopoulos and Patrinos (2018). Formally, we assume $E[(\phi_{q,c}-\phi_{\gamma,c})|\bar{s}_c,x_c] = E[(\phi_{q,c}-\phi_{\gamma,c})|\bar{s}_c]$. If we maintain the assumption of common technology parameters $\gamma(s)$ across countries, as in Caselli and Coleman (2006) and Hendricks and Schoellman (2019), then this reduces to assuming that skill scarcity and school quality are not correlated conditional on years of schooling: $E[\phi_{q,c}|\bar{s}_c,x_c] = E[\phi_{q,c}|\bar{s}_c]$. More generally, conditioning on average years of schooling potentially controls as well for any variation in $\phi_{\gamma,c}$ that might otherwise project on x_c.

The results are shown in Table 4. The first column projects the Mincer return on skill scarcity controlling for a full set of year indicators. The coefficient on scarcity is 0.14 (0.02), implying an elasticity of 7.1 if schooling quality and scarcity are uncorrelated.

The second column controls for average years of schooling as a proxy for school quality.

Two results emerge. First, the Mincer return depends positively on the scarcity of skills with a coefficient of 0.21 (0.04). The implied elasticity of skill substitution is 4.7, which is not far from our estimates in the previous two subsections. Second, conditional on skill scarcity, the coefficient on years of schooling is positive at 0.25 (0.13). This implies that the negative correlation between the Mincer return and years of schooling across countries is indeed explained by differences in skill scarcity.

The working assumption behind these estimates is that the relative slope of parameters governing the frontier of the skill-specific technology, ϕ_γ, is common across countries. It is plausible that ϕ_γ varies across countries. For this to be a concern for the elasticity estimate in Column 2, these relative costs would have to correlate with skill scarcity conditional on average years of schooling. Next, we focus on the trends in the Mincer return and scarcity *within* a country to estimate the elasticity of substitution. This allows for cross-country differences in the levels of relative costs of skill-specific technology, but requires that those differences be relatively stable over time. More exactly, we assume $E[(\Delta\phi_{q,c} - \Delta\phi_{\gamma,c})|\Delta\bar{s}_c, x_c] = E[(\Delta\phi_{q,c} - \Delta\phi_{\gamma,c})|\Delta\bar{s}_c]$, where Δy denotes the time trend of variable y.

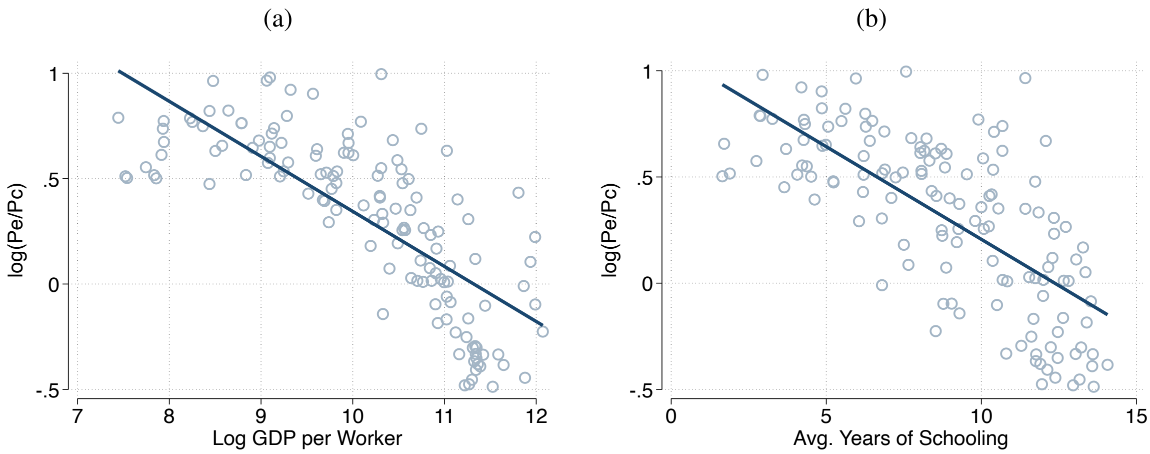
Columns (3) and (4) show the results. Estimating trends in the Mincer return requires multiple years of observations for each country, which reduces the number of countries in our sample from 105 to 83. As skills become less scarce over time, the Mincer return tends to decrease in a country. This decline is sharper when the years of schooling is controlled for in the regression. The estimated elasticity is 5.5 when trends in school quality are ignored, and it is 3.5 when they are not.

3.5 Capital Costs and Skill Bias of Technology

We have shown that plausible differences in school quality cannot reconcile low values of ε with the observed Mincer returns. The only other way to justify a low elasticity of substitution is by technology that is severely biased towards skilled labor in countries where schooling is abundant. To the extent that technology bias is endogenous, as described in 2.2, it will not bias our estimate of the long-run value of the elasticity of substitution. But if poor countries faced steeper costs of skill-specific technology, in addition to having too few skilled workers, then we would expect to see gaps in technical skill-bias beyond what is suggested by schooling endowments. This possibility, represented by higher values of ϕ_γ among countries where schooling is scarce, would bias our estimates upward (see equation (13)).

In our view, the likely effect of variation in ϕ_γ on the estimates above is nonetheless small for two reasons. First, the relative skill-bias required to justify $\varepsilon = 1.5$ is very large. Even with reasonable differences in school quality, the relative cost of skill-specific technology has to be roughly 10 folds in the poor country (25th percentile of GDP per capita) compared to the rich

Figure 6: Relative Price of Equipment Across Countries.



Notes.– Figures shows the relative price of equipment, defined as the log-ratio of price of equipment to the price of consumption, against GDP per capita (Panel a) and average educational attainment (Panel b). Price data comes from the 2011 International Comparison Program of World Bank. Average education is taken from Barro and Lee (2013) and GDP per capita is taken from Penn World Tables 9.1. The latter are 2010 values.

country (75th percentile), assuming similar technology costs for the lowest schooling group. Sustaining such a large gap in the long run demands insurmountable barriers to international flow of technology and capital. Second, differences in skill-specific technology costs are plausibly captured by average educational attainment. Because the estimates of elasticity obtained by controlling for average educational attainment in Section 3.4 are similar to those in Sections 3.2 and 3.3, differences in ϕ_γ cannot be major.

In this subsection, we provide a third reason in support of our view, building on earlier work by (Krusell et al., 2000), who show that capital, particularly machinery and equipment, is more complementary to skilled labor. If equipment is cheaper in rich countries, then skilled labor could be much more efficient than what is suggested by differences in school quality alone. While there is evidence for such complementarity, controlling for cross-country differences in equipment technology does not substantially alter our estimates for the long-run elasticity.

To test for the role of equipment technology, we draw data from the International Comparison Program of the World Bank. The objective of the program is to provide purchasing power parities and comparable price level indices for participating economies. For each country, we compute the relative price of equipment as $\log(P_{eq}/P_{cons})$, where P_{eq} is price index for machinery and equipment and P_{cons} is the price index for household consumption. Both indices are normalized to 100 for the World, resulting in a global aggregate value of zero for the relative price measure.

Figure 6 shows the resulting values by average schooling and labor productivity. There is a strong negative correlation between $\log(P_{eq}/P_{cons})$ and both variables, suggesting that the

Table 5: Estimates of Substitutability: Equipment Prices and Capital-Skill Complementarity

	(1)	(2)
Dependent variable:	$\widehat{m}_c$	$\widehat{m}_c$
$\bar{s}_c$		0.12
		(0.16)
x_c	0.18***	0.21***
	(0.03)	(0.05)
$\log(P_{eq}/P_{cons})_c$	−1.27**	−1.01
	(0.48)	(0.59)
$\widehat{\varepsilon}$	5.60	4.84
Interval	[4.29, 8.04]	[3.32, 8.92]

*** p<0.01, ** p<0.05, * p<0.1

Note.– Table shows the regression of the Mincer return to schooling on average years of schooling, $\bar{s}$, and scarcity of skills, x. The first two columns control for year indicators. Robust standard errors are reported in parentheses.

technology is highly biased towards skilled labor in rich countries, or, in countries with high educational attainment. Relative equipment price in the poor country is about 50 percent higher than it is in the rich country.

We revisit our estimates of ε presented in Table 1 in light of this information on equipment prices. Controlling for the relative price of equipment, the estimates of ε we obtain are 3.8, 4.1 and 4.3. These estimates are slightly lower than 4.1, 4.5 and 4.9 obtained previously and reflect higher cost of skill-specific technology in poor countries. The difference is however not big. Given the revised confidence intervals, values below 2.7 are still rejected. The reason is that the equipment price gap between rich and poor countries, while substantial, are at least an order of magnitude too small to rationalize $\varepsilon = 1.5$.[16]

Next, we revisit Table 4, where we project the Mincer returns on average attainment and schooling. Table 5 shows the results. Without average education among regressors, the estimate of the elasticity of substitution is 5.6 (Column 1). Conditional on scarcity, the Mincer return is lower in countries with higher equipment prices, reflecting gaps in the skill-bias of technology. As a result, the revised estimate of ε is lower than our previous estimate of 7.0.

In Column 2, we include average educational attainment as a regressor, along with equipment price and scarcity. The resulting coefficient on scarcity implies an elasticity of substitution of 4.8, virtually the same as 4.7 we obtained earlier. This suggests that differences in average attainment captures much of the variation in skill-specific technology costs. For confirmation, we projected equipment prices on average attainment and scarcity. The coefficients on scarcity are

[16] We estimate that a country on the 25th percentile of distribution of GDP per capita would need to have roughly 10 times the equipment prices in the rich country, assuming that technology costs associated with the lowest skill group are equal in the two countries. In reality, relative prices are around 50 percent higher in the poor country.

-0.007 (0.006) and 0.03 (0.004) with and without average attainment among regressors. While scarcity and equipment prices are positively correlated in general, they are indeed orthogonal to each other conditional on average attainment.

The upshot of our analysis based on cross-country comparisons of the Mincer return and the scarcity of skills points to an estimate of the elasticity of skill substitution in the 4 to 5 range. Given the standard errors around the estimates, we can generally rule out estimates below 3. Next, we discuss the implications of the elasticity of substitution for long run growth and use accounting methods to infer a plausible range for the elasticity of substitution. This also allows us to relax some of the assumptions we have made in our cross-country analysis.

4 Elasticity of Substitution and Growth Accounting

Schooling attainment has increased greatly in most countries in recent decades. From the Barro-Lee (2013) data, we can calculate years of schooling attainment from 1960 to 2010 for 60 countries for which Mincer returns (from Psacharopoulous and Patrinos, 2018) are observed at least 3 times over the 50 year at intervals of at least five years.[17] As seen from Panel (a) of Figure 7, these countries show an average increase of about six years of schooling. Paralleling this increase in average years has been a tremendous decrease in average scarcity of schooling as measured by x. That is, the increase in schooling has reshaped the distribution of schooling rather than simply shifting it to the right. This is illustrated in Panel (b) of Figure 7.

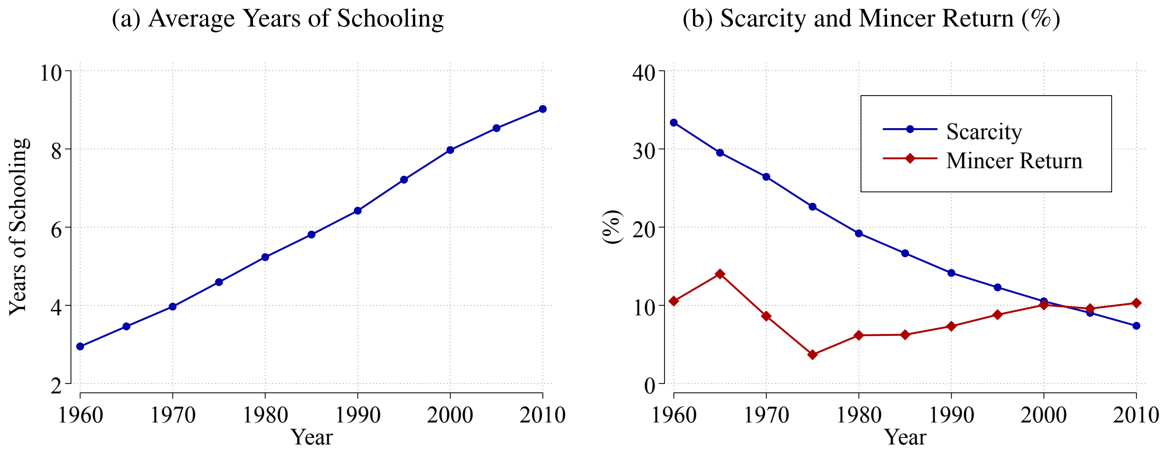
Figure 7, Panel (b) displays the change in average Mincer returns based on the data from Psacharopoulos and Patrinos (2018). For modest substitution across schooling groups, the world-wide decline in x, ceteris paribus, should contribute a substantial decline in the returns to schooling. But, in fact, Mincer returns to schooling have remained largely stable, decreasing from 1965 to 1975, but rebounding ever since.

Explaining these joint trends in scarcity and Mincer returns requires that relative efficiency has grown faster for groups with more schooling, either due to improvements in the quality of schooling or because technological growth has been biased in their favor. If schooling groups are fairly poor substitutes, as typically assumed, this requires spectacularly rapid efficiency gains for those with more schooling. But we show in this section that growth accounting bounds these gains well below such rates given actual rates of growth in real incomes, presuming technology for those with less schooling has not *regressed* dramatically worldwide. In turn, this implies an elasticity of substitution across schooling groups of 4, if not higher.

The human capital aggregator in equation (2) provides a basis for measuring the growth

[17] Countries are weighted by their total employment as of 2000. We drop observations through 1990 for countries that were formerly held in the Soviet Union.

Figure 7: Trends of Educational Attainment, Scarcity, and Mincer Return



Notes.– Data on the Mincer return is taken from Psacharopoulos and Patrinos (2018). Average years of schooling and Scarcity is obtained from authors' calculations based on Barro and Lee (2013). Each trend is obtained by taking average (weighted by total employment in 2000) over the 60 countries in our main sample.

rate in human capital, including the impact from labor augmenting technological change. To that end, equation (12) can be rearranged to express the relative efficiency of each schooling group to that of the lowest group $s_1 = \min\{S\}$ as a function of the ratios of its wage rate and employment to that group.

$$\frac{e(s)}{e(s_1)} = \left[\frac{w(s)}{w(s_1)}\right]^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}} \left[\frac{L(s)}{L(s_1)}\right]^{\frac{1}{\tilde{\varepsilon}-1}}. \tag{17}$$

Substituting this expression in (2) gives:

$$h(\tilde{\varepsilon}) = e(s_1)\left[\frac{\overline{w}}{w(s_1)}\right]^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}} \left[\frac{L}{L(s_1)}\right]^{\frac{1}{\tilde{\varepsilon}-1}} = e(s_1)h_{-e_1}(\tilde{\varepsilon}). \tag{18}$$

Here $\overline{w}$ denotes the average wage in the economy, $\sum_{s\in S}\frac{w(s)L(s)}{L}$. We approximate $\overline{w}$ relative to group 1's wage based on a country's Mincer return: $\frac{\overline{w}}{w(s_1)} = \sum_{s\in S}\left(e^{m(s-s_1)}\frac{L(s)}{L}\right)$.

Equation (18) breaks human capital per worker into two components: the production efficiency of the lowest schooling group, $e(s_1)$, and the quantity of human capital in the economy normalized by that efficiency, $h_{-e_1}(\tilde{\varepsilon})$. $h_{-e_1}(\tilde{\varepsilon})$ increases in both the relative wages and employments of other groups compared to group 1. Dramatic growth worldwide in $\frac{L}{L(s_1)}$, together with the stability of Mincer returns, requires an increase in $h_{-e_1}(\tilde{\varepsilon})$. That increase is especially large if $\tilde{\varepsilon}$ is small. In turn, this implies growth in human capital per worker, h, above and beyond any growth in efficiency of the lowest schooling group, $e(s_1)$.

Efficiency for group s is determined by quality of its schooling, $q(s)$, and its technology, $A(s)$, where the latter reflects both the technology frontier and the choice of technology for s along that frontier. Combining equation (9) for $A(s)$ with (18):

$$e(s) = A(s)q(s) = q(s)\frac{B^{\frac{1}{\omega}}}{\gamma(s)} \cdot \left[\frac{w(s)L(s)}{\overline{w}L}\right]^{\frac{1}{\omega}}. \tag{19}$$

The last term, reflecting group s's relative earnings, captures its directed technology bias along the technology frontier. Substituting in equation (18) for $e(s_1)$ yields:

$$h(\varepsilon) = q(s_1)\frac{B^{\frac{1}{\omega}}}{\gamma(s_1)} \cdot \left[\frac{\overline{w}}{w(s_1)}\right]^{\frac{\varepsilon}{\varepsilon-1}}\left[\frac{L}{L(s_1)}\right]^{\frac{1}{\varepsilon-1}} = z(s_1)h_{-z_1}(\varepsilon) \tag{20}$$

Note two differences from equation (18). $h_{-z_1}(\varepsilon)$ holds $z(s_1)$, not $e(s_1)$, constant. $z(s_1) = q(s_1)\frac{B^{\frac{1}{\omega}}}{\gamma(s_1)}$, is a combination of school quality for the minimal schooling group and the technology frontier as reflected in B and $\gamma(s_1)$. Thus $z(s_1)$ allows for a response of technology across groups to their relative supplies. The second difference is that it is the long-run elasticity ε that governs how $h_{-z_1}(\varepsilon)$ responds to relative wages and employments across groups.

In turn, the growth rate in human capital per worker, defined in log-differences, can be viewed in light of the growth rate in these two terms:

$$g_h(\varepsilon) = g_{z_1} + g_{h-z_1}(\varepsilon), \tag{21}$$

where g_{z_1} denotes the growth rate of z_1 and $g_{h-z_1}(\varepsilon)$ that of $h_{-z_1}(\varepsilon)$. For any assumed elasticity ε, $g_{h-z_1}(\varepsilon)$ can be calculated from how the schooling distribution of workers and Mincer return evolve over time. Our strategy is to construct an implied lower bound for $g_h(\varepsilon)$ under various values for ε by: (a) measuring $g_{-z_1}(\varepsilon)$ from cross-country data for each ε, and (b) assuming a plausible lower bound for g_{z_1}, that is, the growth in efficiency of the lowest group reflected by changes in its schooling quality and growth in the technology frontier.

We employ $g_{z_1} = 0$ for its lower bound. We view this as conservative. For instance, if one assumes no change in schooling quality for group s_1, this would require no improvement on average from the technological frontier for these workers world-wide since 1960. Note this bound does allow substantial technological regress for group s_1 in the form of *directed* technological change: As highlighted in (19), that directed change will be away from group s_1 given the large secular decline in its relative earnings share.[18] We treat those with completed primary education or less as the group of minimal schooling. While it is imaginable that schooling

[18]If one alternatively assume zero growth in $e(s_1)$ over time, that would generally imply a less conservative lower bound. From equation (18), this yields $g_h(\tilde{\varepsilon}) = g_{e_1} + g_{-e_1}(\tilde{\varepsilon})$. Assuming $g_{e_1} = 0$ here yields the same $g_h(\tilde{\varepsilon})$ as assuming $g_{z_1} = 0$ does for $g_h(\varepsilon)$. Thus this alternative bound would yield identical implications for $\tilde{\varepsilon} \leq \varepsilon$ that we report for ε.

quality has declined over time, we would not anticipate a decline in quality of primary schooling worldwide since 1960.[19] But, regardless, given the limited schooling received for these workers, any such quality decline should be swamped in importance by worldwide gains in the technology frontier for these workers, especially recognizing that this group averaged about 40% of the working age population for these countries as of 2000. Therefore, we view $g_{z_1} = 0$ as providing a conservative lower bound for $g_h(\varepsilon)$.

Estimation in Sections 3.2 to 3.4 assumed technology parameters across countries, $\gamma(s)$, such that ϕ_γ did not project on scarcity in a country, x, at least not conditioning on a country's average schooling attainment. Note that assuming that *on average* $g_{z_1} = 0$ puts no such constraint on the pattern of ϕ_γ across countries. It only requires that on average worldwide there was not technological regress in the technology frontier, combined with school quality, for those with the lowest schooling level.

Alternatively g_h can be measured from standard growth accounting, given an economy's growth rates in output and capital. From the aggregate technology in (1), output per worker, $y = Y/\sum_s L(s)$, is given by $y = k^\alpha h^{1-\alpha}$ where k, like h, denotes input per worker. The growth rate of human capital is:

$$g_h = \frac{1}{1-\alpha}(g_y - \alpha g_k), \tag{22}$$

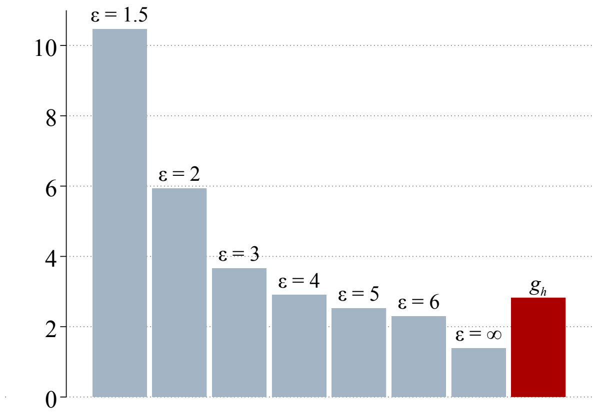
where, again, g_h, reflects the impact of technical change as well as increases from schooling investments. Thus, by comparing $g_h(\varepsilon)$ from (21) to its estimate from growth accounting we can judge plausible magnitudes for ε.

In Figure 8, we contrast the rate of human capital growth between 1960 and 2010 implied by the two accounting methods for 60 countries for which we have estimated Mincer returns for at least three points in time.[20] The data on output and capital per worker behind g_h come from the Penn World Tables 9.1 (Feenstra, Inklaar, and Timmer, 2015). For $g_h(\varepsilon)$, data on educational attainment for the working age population, 25 to 54, are from Barro and Lee (2013). The Mincer returns are taken from Psacharopoulos and Patrinos (2018). The black bar to the right shows the growth rate g_h from the growth accounting equation (22) assuming that α, capital's share in output, is $1/3$. The gray bars show lower bounds for $g_h(\varepsilon)$ implied by the changes in schooling distributions and Mincer return under differing values of the elasticity of substitution. Each gray bar presumes zero average growth worldwide from 1960 to 2010 for the lowest schooling group from school quality and the technology frontier ($g_{z_1} = 0$), where we

[19]Compositional changes could affect $q(s_1)$, as we discuss below in the context of the findings.

[20]Countries are weighted by total employment as of 2000. The average begin and ending years across the countries are 1973 and 2007. For most countries the share of the bottom schooling group, completed primary or less, averages more than 20 percent across the years. It is considerably lower for the four countries: Canada, Austria, United States, and Latvia. But dropping these four countries does not appreciably affect Figure 8 or the implied value for ε discussed just below.

Figure 8: Elasticity of Skill Substitution and the Implied Rate of Human Capital Growth (%)



Note.– Figure depicts the average annual growth rate of human capital for the 60 countries in our sample. The gray bars show the mean $g_h(\varepsilon)$ implied by different values of the elasticity of skill substitution (ε), using the changing distributions of schooling and Mincer returns. The red bar shows $g_h(\varepsilon)$ from growth accounting given growth in per worker income and capital. Countries are weighted by their relative GDPs for 2000. Source: Authors' calculations based on Psacharopoulos and Patrinos (2018), Barro and Lee (2013) and Penn World Tables 9.1.

treat completed primary or less as the minimal schooling group,

$g_h(\varepsilon)$ is strongly decreasing in the assumed elasticity of substitution. The reasoning is straightforward. The worldwide decrease in scarcity of higher schooling groups was a force to reduce Mincer returns to schooling, especially if the elasticity of substitution is low. Because Mincer returns were essentially stable despite declining scarcity, higher schooling groups must have become more efficient over time. For low elasticities, that rate of efficiency gain must be extremely rapid. In particular, for $\varepsilon = 1.5$, $g_h(\varepsilon)$ must average 10.4% per year during the 50 years for the 60 countries.

But such rapid growth in human capital is sharply at odds with actual output growth around the world: A 10.4% growth rate in human capital, given a labor share of two-thirds, produces annual output growth of 7.0% even neglecting capital's growth; that far exceeds actual rates. The dark bar in Figure 8 shows that the rate of human capital growth that is consistent with the observed output growth is instead 2.8%. Assuming, on average, no growth in schooling quality or in the technology frontier for the lowest schooling group (average $g_{z_1} = 0$), an elasticity of substitution of 4.15 is required to reconcile the growth rate of human capital from the constructive accounting, $g_h(\varepsilon)$, with that from growth accounting.[21]

If we allow for improvements in the technology frontier or school quality for the lowest

[21] We can alternatively choose the value for ε that minimizes the distance between the two measures, g_h and $g_h(\varepsilon)$ across our sample of 60 countries. That exercise weighs matching the correlation in the two measures across the countries as well as their means. Formally, we numerically solve the problem: $\varepsilon^* = \min_\varepsilon \sum_c \left[g_{h,c} - g_{-1,c}(\varepsilon)\right]^2$, where c denotes a country. This yields a $\varepsilon^* = 5.48$, with 95% confidence interval [3.28, 129.82].

schooling group, then $g_h(\varepsilon)$ is directly increased by $g_{z_1} > 0$. This implies the required elasticity must be adjusted *upward* to maintain consistency with the data. While technological regress is unlikely, it is conceivable that average skills among the least educated workers fall over time if they become selected negatively on other dimensions of productivity, such as ability. In that case, the estimated elasticity above has to be adjusted downward to reflect the difference between the evolving selection effect and the rate of technical progress. Adjusting for plausible selection, however, does not substantially alter our conclusions.[22]

5 Robustness

In this section, we investigate the sensitivity of our estimates along a couple of dimensions. First, we consider alternative groupings of the population into schooling categories. Second, we examine if substituting schooling groups by population for groupings by employment, as we have in the previous sections, biases our estimates. In each case, the upshot of our analysis is that the estimates reported above do not change in any significant way.

5.1 Grouping of School Categories

In Sections 3 and 4, we divide labor into four imperfectly substitutable schooling categories: i) complete primary and below, ii) some secondary, iii) complete secondary, and iv) some tertiary and above. In this section, we examine the sensitivity of our results with respect to alternative groupings of Barro and Lee's schooling data. Specifically, we consider the following cases:[23]

(a) 2 Groups-1: complete secondary and below; some tertiary and above.

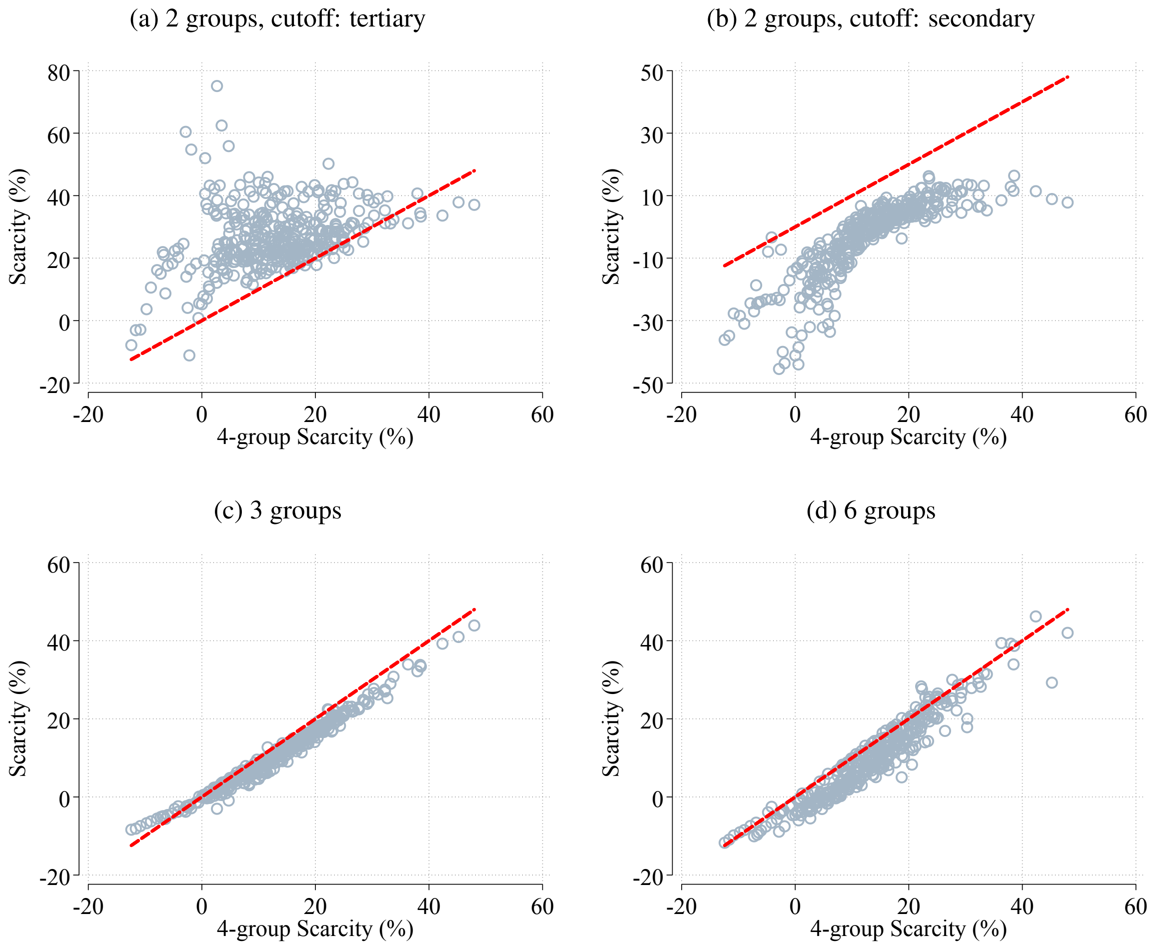
(b) 2 Groups-2: complete primary and below; some secondary and above.

(c) 3 Groups: complete primary and below; some or complete secondary; some tertiary and above.

[22] In particular, to rationalize a value for ε of 1.5 requires this selection reduces the quality of the bottom group by 7.6% per year. Given the rate of worker outflow from the bottom group, this would require that those reductions fell *on average* on workers who have 3.6 times the group's average productivity. This is an implausible differential. Given average Mincer returns across countries (from Schoellman (2012)) this is equivalent to the earnings difference associated with 26 years of additional schooling. To justify values for ε of 2 or 3, still requires implausible selection. $\varepsilon = 2$ requires those exiting exhibit a productivity differential comparable to 14 years of schooling; for $\varepsilon = 3$ it would be 9 years of schooling. Even the latter implies that those workers selecting out of the bottom group are equivalent to workers with some college, despite having only a primary education.

[23] We also considered two case with five groupings, one that departs from the benchmark four groups by dividing the lowest schooling level between partial and completed primary, the other that departs by dividing the highest level between partial and complete tertiary. Both cases produce results similar to those under 3, 4, or 6 groups, with $\widehat{\varepsilon}$ varying from 3.8 to 6.3 across the specifications reported in Table 6.

Figure 9: Scarcity with differing number of groups versus Benchmark 4 group Scarcity.



Note.– Red dash lines depict the 45° lines.

(d) 6 Groups: some primary and below; complete primary; some secondary; complete secondary; some tertiary; complete tertiary.

Within each category, the corresponding years of schooling is obtained by taking the population weighted average from the original groups in the data set.

Figure 9 compares the schooling scarcity calculated under each grouping rule with our benchmark grouping for the main sample used in Section 3 (105 countries and 371 observations). The red, dashed lines depict the 45° lines. The groupings by three and six levels, Panels c and d, measure scarcity similarly to the benchmark grouping of four. The groupings of two, Panels a and b, both diverge from the benchmark measure but in different directions depending on the choice for the cutoff. Generally, grouping by fewer categories implies a bigger share of variations in schooling are manifested within-group, muting the variation in measured scarcity either across countries or over time. This is especially true for the 2-group cases, and most notably when the cutoff is further away from the median schooling level as in Panel (a). Because grouping more sparsely results in a loss of valuable information for identifying the elasticity of substitution, we can expect these to provide less precise estimates for ε. Furthermore, to the extent these provide a cruder measure of scarcity, we can expect they will yield a lower estimate for $1/\varepsilon$; so, therefore, a higher estimate for ε. The takeaway from Figure 9 is that distinctions between primary and secondary and between secondary and tertiary schooling are both important components of scarcity; but groupings of two treat those nontrivial features of the schooling distribution as irrelevant for the Mincer return. At the minimum, a separation between primary, secondary, and tertiary schooling is needed.

Table 6 reinforces this observation. We repeat the estimation in Section 3 under each grouping rules. Recall that the four-group case is our benchmark (fourth column). The last three columns of all panels shows that our main conclusion is robust for three-group cases or above. Estimates for the elasticity of substitution are generally located between 3.7 and 5.8, and elasticities smaller than 2.5 are clearly rejected. The first two columns, however, are quite different from our benchmark results. Grouping into two-categories in general leads to much larger estimated elasticities. Based on the confidence intervals, however, we find these estimates to be generally unreliable.

Panel D, which revisits the estimates in Section 3.4, highlights the identification problem faced by groupings of two. Recall that the Section 3.4 approach is to use average schooling attainment to account for cross-country differences in school quality. Because average schooling is a sufficient statistic for scarcity when there are two groups, no independent variation remains (i.e., from skewness) to estimate the elasticity of substitution.

5.2 Schooling Distribution by Employment

In our calculations of scarcity, we rely on the population shares of schooling from Barro and Lee (2013), whereas the wage equation in (12) stipulates the relative shares of schooling in the work force. If differences in employment rates by schooling are systematically linked to scarcity of schooling, then the cross-country estimates in Section 3 could be biased. To gauge the severity of this bias, we draw data on employment and population by schooling levels from the International Labor Organisation (ILO). The sample covers 121 countries for the years 1990 to 2018, though data on most countries begin after 2002.[24] We compute scarcity of schooling both in the population and among workers and plot them in Figure 10. The dashed line shows the 45-degree line. The two measures are correlated almost perfectly, with a coefficient of 0.98. This suggests that any bias introduced by substituting population scarcity for scarcity among workers cannot be significant.

[24] We use population-based measures of schooling scarcity in Section 3 despite the availability of employment measures in the ILO, because the ILO data is too recent relative to the data on Mincer returns in Psacharopoulos and Patrinos (2018), and, as a result, there is little overlap between the two data sets.

Table 6: Estimates of Substitutability with Different Grouping

Panel A: OLS

	2 Groups (1)	2 Groups (2)	3 Groups	4 Groups	6 Groups
Slope	0.10	0.12	0.27	0.25	0.22
	(0.04)	(0.03)	(0.06)	(0.05)	(0.05)
$\widehat{\varepsilon}$	10.16	8.14	3.69	4.05	4.51
Interval	[5.42, 80.84]	[5.29, 17.62]	[2.54, 6.77]	[2.81, 7.27]	[3.07, 8.41]

Panel B: GMM (IV = x_c)

	2 Groups (1)	2 Groups (2)	3 Groups	4 Groups	6 Groups
Slope	0.05	0.10	0.22	0.24	0.20
	(0.04)	(0.03)	(0.06)	(0.05)	(0.06)
$\widehat{\varepsilon}$	19.89	9.66	4.17	4.45	5.04
Interval	[7.57, ∞]	[5.85, 27.69]	[2.73 8.87]	[2.99 8.69]	[3.20 11.95]

Panel C: GMM (IV = y_c, $\bar{s}_c$)

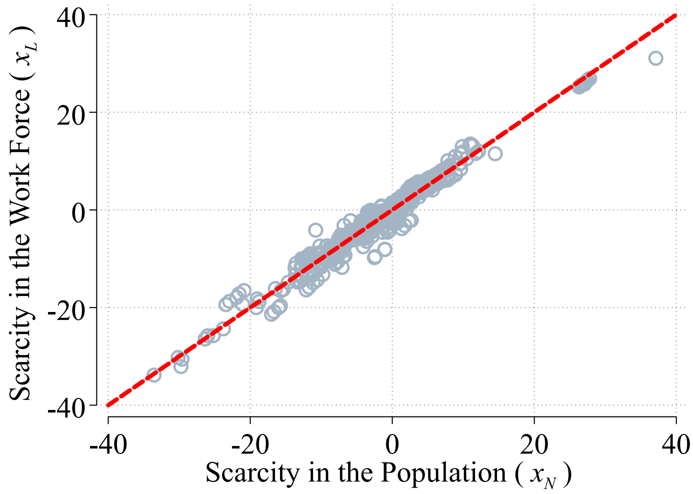
	2 Groups (1)	2 Groups (2)	3 Groups	4 Groups	6 Groups
Slope	0.29	0.12	0.24	0.20	0.17
	(0.15)	(0.04)	(0.08)	(0.07)	(0.06)
$\widehat{\varepsilon}$	3.47	8.27	4.21	4.94	5.80
Interval	[1.68, ∞]	[4.76, 31.76]	[2.55, 12.19]	[2.98, 14.43]	[3.51, 16.78]

Panel D: OLS ($\widehat{m}_c$ on $\bar{s}_c$ and x_c)

	2 Groups (1)	2 Groups (2)	3 Groups	4 Groups	6 Groups
Slope	–	–	0.20	0.21	0.25
	–	–	(0.04)	(0.05)	(0.06)
$\widehat{\varepsilon}$	–	–	5.05	4.66	3.93
Interval	–	–	[3.51, 8.95]	[3.28, 8.05]	[2.68, 7.35]

Note.– Each Panel reports the estimation results under different grouping rules. The first 3 panels reproduce the analyses in Section 3.2, regressing $(\widehat{m} - \widehat{\phi}_q)$ on $(x - \widehat{\phi}_q)$ across 51 countries in 2000. Panel A shows the OLS estimates, and Panel B and C shows the GMM estimates with different instrument variables. Panel D reproduce the analyses in Section 3.4, reporting OLS regression of $\widehat{m}$ on x controlling for $\bar{s}$ and year indicators using the sample with 105 countries and 371 observations.

Figure 10: Scarcity by Population and Employment



Note.– Red dash line depicts the 45° line. Source: International Labor Organisation..

We revisit three specifications from Section 3. To gauge the potential biases in our estimates of the elasticity of substitution, we project the employment-based measures of scarcity on the population-based measures. The regression coefficient indicates the relative bias in the estimation of ε. For instance, in Section 3.2, we used immigrants' return to schooling to control for differences in schooling quality across countries. Our benchmark specification in Column 1 of Table 1 involved projecting $\widehat{m} - \widehat{m}^{US}$ on $x_N - \widehat{m}^{US}$ to recover $1/\varepsilon$, because the corresponding data on $x_L - \widehat{m}^{US}$ were not available. Denote the relation between the two scarcity measures with the following projection:

$$x_L - \widehat{m}^{US} = \lambda_0 + \lambda_x(x_N - \widehat{m}^{US}) + e$$

By substituting the population-based measure, we would have estimated λ_x/ε instead of $1/\varepsilon$. If $\lambda_x > 1$, then our estimate of ε is biased downward. We estimate λ_x in the ILO data where both measures of scarcity are readily available, and update our estimates accordingly.

Table 7 shows the results for three specifications. The first column revisits the estimate in Column 1 of Table 1. The coefficient of 1.04 (0.01) implies that the elasticity of 4.05 obtained there was biased downward by 4 percent. The corrected estimate is 4.2.[25] Columns 2 and 3 in Table 7 estimate the potential biases in the estimates presented in Table 4, where we used average educational attainment to account for the cross-country variation in technical efficiency of schooling groups. The estimates in both cases are around one, implying that substituting population-based measures of scarcity does not generate a bias in our estimates.

[25] If, $\widehat{m}^{US}$ is measured with error, then the estimate of λ_x is biased towards one, implying that the true elasticity of substitution is likely even larger.

Table 7: Sensitivity of Estimates to using Population Scarcity

	(1) $x_L - \widehat{m}^{US}$	(2) x_L	(3) x_L
$x_N - \widehat{m}^{US}$	1.04***		1.03***
	(0.01)		(0.01)
x_N		1.00***	
		(0.01)	
$\bar{s}_c/100$			–0.23***
			(0.02)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note.– Table shows the projection of schooling scarcity among workers on the scarcity in the population. All specifications include year fixed effects. Sample covers 121 countries between 1990 and 2018. Data comes from the ILO.

6 Conclusion

Starting from production that aggregates workers of differing schooling groups, we outline how an economy's Mincerian return reflects quality of its schooling, skill-bias of its technology, and scarcity of its more-educated workers. We show that country scarcity relevant for the Mincer return, given a CES aggregator, reflects the product of its mean schooling and, essentially, skewness of its schooling distribution. We exploit this measure to gauge the long-run elasticity of substitution between groups, where that elasticity allows for technology to respond to relative earnings across groups, as in Caselli and Coleman (2006), Acemoglu (2007), and Hendricks and Schoellman (2019).

Workers with more schooling are much less scarce in richer countries, yet Mincer returns are nearly as high in those countries. For smaller elasticities, this requires that schooling generates enormously greater gains in efficiency in richer countries. For instance, for an elasticity of 1.5, even in the worst case where schooling brings no efficiency gains in the poorest countries, a *each year* of schooling would have to yield about a 60% increase in labor efficiency in richest countries. We do not see anything near such efficiency gains in the two measures of school quality we examine based on earnings of immigrants to the U.S. and based on standardized test scores. These proxies are instead consistent with a long-run elasticity of about 4.

Finally, and perhaps most telling, we show that growth accounting points strongly to values of ε of 4 or above. Lower long-run elasticities, especially those of two or less, imply rapid technological regress for a large section of the workforce world-wide for 1960 to 2010, even beyond that from technology shifting endogenously towards workers with more schooling.

The elasticity of substitution plays an important role in several quantitative literatures. It is obviously important for understanding the evolution of earnings inequality.[26] In recent years, it has played a central role in papers examining the role of human capital in income differences across countries (see Jones, 2014; Caselli, 2016; Caselli and Ciccone, 2019, among others).

Early papers by Klenow and Rodrigues-Clare (1997) and Hall and Jones (1999) treated schooling groups as perfect, but unequal, substitutes. Both find richer countries have significantly higher human capital per worker; but those differences loom small relative to the enormous differences in worker productivities across countries. Jones (2014), Caselli (2016), and Malmberg (2018) each entertain finite elasticities, with the some focus on elasticities on the order of 1.5, given the estimates from the literature. This implies greater differences in efficiencies across schooling types in rich versus poorer countries, much greater for $\varepsilon = 1.5$. In turn, this yields much bigger differences in average efficiencies between rich and poorer countries assuming that less educated workers in rich countries are not less efficient than they are in poor countries. Jones (2014) goes further to illustrate that for ε's near 1.5, replacing human capital efficiencies in a representative poor country (15th decile of income per worker) with those of workers in a rich country (85th decile) would eliminate about 100 percent of the differences in income per worker between those countries.[27]

Our estimates for the elasticity ε imply much smaller differences in efficiencies across workers in rich countries, compared to poor. For an exercise such as Jones', it would imply differences in worker efficiencies only about one-fifth that implied by $\varepsilon = 1.5$, while about twice that under perfect substitutes ($\varepsilon = \infty$). In fact, even values for ε around 2 would cut the differences in efficiencies by something like two-thirds.

References

Acemoglu, Daron. 2007. "Equilibrium bias of technology." *Econometrica* 75 (5):1371–1409.

Acemoglu, Daron and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of labor economics*, vol. 4. Elsevier, 1043–1171.

Altonji, J. and C. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1):313–350.

[26] Ciccone and Peri (2005) review much of that work.

[27] Caselli and Ciccone (2019) point out the important subtlety in interpreting this exercise. If the higher relative productivity of workers in richer countries reflects technology bias toward skilled workers, then it is not clear one should view this as a human capital accounting. In fact, even if it reflects quality differences in schooling, that might be better labeled as a contribution of technology – if a country had a more productive capital-goods sector, we would not typically attribute that contribution to income as capital deepening.

Barro, Robert J. and Jong Wha Lee. 2013. “A New Data Set of Educational Attainment in the World, 1950–2010.” *Journal of Development Economics* .

Card, David and Thomas Lemieux. 2001. “Can falling supply explain the rising return to college for younger men? A cohort-based analysis.” *The Quarterly Journal of Economics* 116 (2):705–746.

Caselli, Fracesco. 2016. *Technology Differences over Space and Time (CREI Lectures in Macroeconomics)*. Princeton University Press.

Caselli, Francesco and Antonio Ciccone. 2013. “The Contribution of Schooling in Development Accounting: Results from a Nonparametric Upper Bound.” *Journal of Development Economics* 104 (199-211).

———. 2019. “The Human Capital Stock: A Generalized Approach: Comment.” *American Economic Review* 109 (3):1155–1174.

Caselli, Francesco and Wilbur John Coleman. 2006. “The World Technology Frontier.” *American Economic Review* 96 (3):499–522.

Ciccone, Antonio and Giovanni Peri. 2005. “Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950 –1990.” *The Review of Economics and Statistics* 87 (4):652–663.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. “The Next Generation of the Penn World Table.” *American Economic Review* .

Feldstein, Martin S. 1973. “On the Optimal Progressivity of the Income Tax.” *Journal of Public Economics* 2 (4):357–376.

Hall, Robert E. and Charles I. Jones. 1999. “Why Do Some Countries Produce So Much More Ouput Per Worker Than Others?” *Quarterly Journal of Economics* 114 (1):83–116.

Heckman, James J, Lance Lochner, and Christopher Taber. 1998. “Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents.” *Review of Economic Dynamics* 1 (1):1–58.

Hendricks, Lutz and Todd Schoellman. 2018. “Human Capital and Development Accounting: New Evidence from Wage Gains at Migration.” *Quarterly Journal of Economics* 133 (2):665–700.

———. 2019. “Skilled Labor Productivity and Cross-country Income Differences.” *Working Paper* .

Johnson, Matthew and Michael P Keane. 2013. “A dynamic equilibrium model of the US wage structure, 1968–1996.” *Journal of Labor Economics* 31 (1):1–49.

Jones, Benjamin F. 2014. “The Human Capital Stock: A Generalized Approach.” *American Economic Review* .

Katz, Lawrence F. and Kelvin M. Murphy. 1992. “Changes in Relative Wages, 1963-1987: Supply and Demand Factors.” *Quarterly Journal of Economics* 107 (1):35–78.

Klenow, Pete and Andres Rodrigues-Clare. 1997. “The Neoclassical REvival in Growh Economics: Has It Gone Too Far?” In *NBER Macroeconomics Annual*, vol. 12, edited by Ben S. Bernanke and Julio Rotemberg. MIT Press, 73–114.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante. 2000. “Capital-skill complementarity and inequality: A macroeconomic analysis.” *Econometrica* 68 (5):1029–1053.

Malmberg, Hannes. 2018. “How Does the Efficiency of Skilled Labor Vary across Rich and Poor Countries? An Analysis Using Trade and Industry Data.” *Working Paper* .

Psacharopoulos, George and Harry Antony Patrinos. 2018. “Returns to Investment in Education: A Decennial Review of the Global Literature.” *Education Economics* .

Schoellman, Todd. 2012. “Education Quality and Development Accounting.” *Review of Economic Studies* 79:388–417.

Tinbergen, Jan. 1975. *Income Distribution: Analysis and Policies*. New York: North-Holland Publishing Company.

Appendix

A Data Appendix

Sections 3 through 5 make use of cross-country panel data on schooling attainments and estimated Mincer returns. The data on educational attainment by country are from Barro and Lee (2013). See http://www.barrolee.com/. The data include 153 countries with attainments reported at five-year intervals from 1950 to 2010. It contains population frequency distributions over 7 educational categories by broad age groups. We restrict our population sample to those ages 25 to 54. We associate a number of years of schooling to each attainment category using UNESCO Institute for Statistics (http://data.uis.unesco.org/) data on the duration of educational categories for each country Our benchmark case divides workers into four groups: i) completed primary or less, ii) some secondary schooling, iii) completed secondary, and iv) at least some tertiary. For our 105 country sample for Section 3, the average shares by group are respectively 47%, 20%, 22%, and 11%.

The data on the Mincer return are obtained from Psacharopoulos and Patrinos (2018). The paper compiles 1,120 estimates of Mincer wage equations, from micro data on workers' wages, ages and education, for 139 countries going back before 1960. In cases where multiple Mincer estimates are available for a country at the same 5-year interval, we use the average of those estimates. We merge the Barro-Lee and Psacharopoulos and Patrinos (2018) data to obtain an unbalanced panel sample of 371 observations for 105 countries spanning years 1960 to 2010 on both attainment and and the Mincer return to schooling.

In Section 3 we often stratify countries by their average years of schooling attainment or by their Log real GDP per worker. Attainment is from Barro and Lee (2013), as just discussed. Real GDP per worker is obtained from Penn World Tables 9.1. (See https://www.rug.nl/ggdc/productivity/pwt/?lang=en.)

We consider two measures of schooling quality across countries. In Section 3.2 we employ Schoellman (2012)'s estimates of a country's schooling quality based on U.S. earnings of immigrants who received all or most of their schooling in their country of birth. See the supplementary data cited by Schoellman (2012) in his Table A1. There are 51 countries from Schoellman's estimates for which we are have estimates of schooling attainment and Mincer schooling returns estimated on earnings in the home countries.

Our second measure is based on standardized test scores across countries, more precisely on the gradient of the test score with respect to years of schooling by country. The testing is overseen by the Programme for International Student Assessment (PISA). These tests are given to age 15 students on three areas: mathematics, science and reading. We construct two school

Table 8: List of Countries in Sample

Country Name				
Albania†	**Cyprus**	Iraq	**Netherlands**†	**Spain***†
Algeria*†	Czech Republic*†	Ireland†	New Zealand†	**Sri Lanka***
Argentina*	**Denmark**†	**Israel***†	**Nicaragua***	**Sudan***
Australia*†	Dominican Rep.†	**Italy***†	Niger	**Sweden**†
Austria†	**Ecuador***	Jamaica	**Norway**†	**Switzerland***†
Bangladesh*	**Egypt***	**Japan**†	**Pakistan***	**Taiwan***
Belgium†	El Salvador	Jordan*†	**Panama***	Tajikistan
Belize	Estonia†	Kazakhstan	Papua New Guinea	**Tanzania**
Bolivia*	**Finland**†	**Kenya**	Paraguay	**Thailand***†
Botswana	**France***†	Kuwait	**Peru***†	**Tunisia**†
Brazil*†	Gambia	Kyrgyzstan	**Philippines***	**Turkey***†
Bulgaria*†	**Germany**†	**Latvia***†	**Poland***†	**Uganda**
Cambodia	**Ghana***	Malawi	**Portugal***†	Ukraine*
Cameroon	**Greece***†	**Malaysia***	**Romania**†	United Arab Emirates†
Canada*†	**Guatemala***	Maldives	Russian†	**United Kingdom***†
Chile†	Honduras	Malta†	Rwanda	**USA***†
China*	Hong Kong	**Mexico***†	Singapore*†	Uruguay
Colombia*†	**Hungary***†	Mongolia	Slovakia*†	**Venezuela***
Costa Rica	**India**	Morocco*	**Slovenia**†	**Viet Nam***†
Cote dIvoire	**Indonesia**†	Namibia	**South Africa***	Zambia
Croatia*†	**Iran***	Nepal*	**South Korea***†	Zimbabwe

Note.– Table lists the 105 countries whose Mincer return and schooling distribution are observed at least once between 1960-2010. The marker "∗" indicates countries that are used for regression analysis in Section 3.2 (Table 1), and "†" indicates those for Section 3.3 (Table 3). The countries in bold are used for the growth accounting analysis in Section 4. The ordering of country names is alphabetical.

quality measures based on the micro-level data from the 2015 wave of test, as discussed in Section (3.3). These data are available from the OECD. (See https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf.) We map these these test scores to their implications for wages based on the relationship between wage rates and a standardized test score, Armed Forces Qualification Tests, we estimate from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY). (See https://www.bls.gov/nls/nlsy79.htm.) Our estimates are based on men with at least 15 years of work experience.

For growth accounting in Section 4, we require data on real output per worker and real capital input per worker. These are constructed from version 9.1 of the Penn World Tables referenced above.

Table 8 lists the countries reflected in the empirical work, denoting each exercise that a country could be utilized.

B Scarcity of Schooling: Average Schooling versus Skewness

This section derives the relation between our scarcity index $x = Cov[\ln l(s), s]/Var(s)$ and the mean and skewness of a schooling distribution.

But, first consider the case of only two-groups, distributed with a fixed support $S = \{s_1, s_2\}$. Given mean attainment, $\bar{s}$, the shares of each group and scarcity are determined: $l(s_1) = (s_2 - \bar{s})/(s_2 - s_1)$, $l(s_2) = (\bar{s} - s_1)/(s_2 - s_1)$, and

$$x = -\frac{\ln l(s_2) - \ln l(s_1)}{s_2 - s_1} = \ln\left(\frac{s_2 - \bar{s}}{\bar{s} - s_1}\right)/(s_2 - s_1). \tag{23}$$

So, with only two groups, there is no variation in x conditional on $\bar{s}$; and x decreases with $\bar{s}$. In the cross-country data, the schooling years s_1 and s_2 are calculated by taking the average schooling with the group, hence they varies across countries and times.

More generally $\bar{s}$ does not imply scarcity x. What does the variation in scarcity capture conditional on average schooling attainment? Let $l(s) = L(s)/\sum L(s)$ be the probability distribution of schooling. The scarcity index can be expressed as:

$$\begin{aligned}
x &= -Cov\big[\ln l(s), s\big]/\sigma_s^2 \\
&= -\frac{1}{\sigma_s^2}\Big\{\sum_S \Big(\ln l(s)\Big) s l(s) - \Big[\sum_S \Big(\ln l(s)\Big) l(s)\Big] \cdot \Big[\sum_{s \in S} s l(s)\Big]\Big\} \\
&= \frac{-1}{\sigma_s^2}\Big\{\sum_S \Big(\ln l(s)\Big) \frac{s l(s)}{\bar{s}} - \Big[\sum_S \Big(\ln l(s)\Big) l(s)\Big]\Big\} \cdot \bar{s} \\
&= \frac{-1}{\sigma_s^2}\Big[\sum_S \Big(\ln l(s)\Big) l(s) \Big(\frac{s - \bar{s}}{\bar{s}}\Big)\Big] \cdot \bar{s}.
\end{aligned} \tag{24}$$

Now define

$$\kappa = \frac{-1}{\sigma_s^2}\Big[\sum_S \Big(\ln l(s)\Big) l(s) \Big(\frac{s - \bar{s}}{\bar{s}}\Big)\Big]. \tag{25}$$

so the scarcity can be written as $x = \kappa \cdot \bar{s}$. We want to express κ as in equation (6), hence showing that it measures the asymmetry (skewness) of the schooling distribution.

Note that the term within the summation is positive for all $s > \bar{s}$ and is negative for all $s < \bar{s}$. So we can separate the summation into two parts.

$$\kappa = \frac{1}{\sigma_s^2}\Big[\sum_{s<\bar{s}} \Big(\ln l(s)\Big) l(s) \Big|\frac{s - \bar{s}}{\bar{s}}\Big| - \sum_{s \geq \bar{s}} \Big(\ln l(s)\Big) l(s) \Big|\frac{s - \bar{s}}{\bar{s}}\Big|\Big]. \tag{26}$$

In addition, $\sum_S l(s)\left(\frac{s-\bar{s}}{\bar{s}}\right) = 0$, giving us:

$$\sum_{s<\bar{s}} l(s)\left|\frac{s-\bar{s}}{\bar{s}}\right| = \sum_{s\geq\bar{s}} l(s)\left|\frac{s-\bar{s}}{\bar{s}}\right| = \frac{1}{2}\sum_{S} l(s)\left|\frac{s-\bar{s}}{\bar{s}}\right|. \tag{27}$$

Equation (26) can thus be rewritten as:

$$\kappa = \frac{\sum_S \left(l(s)|\frac{s-\bar{s}}{\bar{s}}|\right)}{2\sigma_s^2}\left[\mathbb{E}_{l(s)|\frac{s-\bar{s}}{\bar{s}}|}\left(\ln l(s)\middle|s<\bar{s}\right) - \mathbb{E}_{l(s)|\frac{s-\bar{s}}{\bar{s}}|}\left(\ln l(s)\middle|s\geq\bar{s}\right)\right], \tag{28}$$

Yielding equation (6).

C Deriving the Long-run Elasticity of Substitution

As in, Caselli and Coleman (2006) we consider an economy with a large number of competitive firms, with labor and capital supplied elastically. The representative firm solves the optimization problem:

$$\max_{\{L(s),A(s)\},K} K^{\alpha}H^{1-\alpha} - \sum_{s\in S} w(s)L(s) - RK, \tag{29}$$

subject to the technological frontier

$$\sum_{s\in S}\left[\gamma(s)A(s)\right]^{\omega} \leq B, \tag{30}$$

and where effective human capital, H, aggregates labor over skill groups

$$H = \left[\sum_{s\in S}\left(A(s)q(s)L(s)\right)^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}}\right]^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}}. \tag{31}$$

An equilibrium consists of prices $(\{w(s)\}_s, R)$ and allocations $(\{L(s),A(s)\}_s, K)$ such that input markets clear subject to firms' optimized at those prices.

We next show the condition for a symmetric equilibrium with interior solution. It enables us to characterize the equilibrium by the first-order conditions of a representative firm. Then we derive the long-run elasticity of substitution, which parallels Hendricks and Schoellman (2019)'s treatment.

C.1 Symmetric Equilibrium with Interior Solution

We want to show that $\omega - \tilde{\varepsilon} + 1 > 0$ is a sufficient condition for a symmetric equilibrium with interior solution. A symmetric equilibrium means all firms chose the same technology bundles, and interior solution means $A(s) > 0$ for all $s \in S$.

First we make the change of variables $D(s) = A(s)^{\omega}$ and rewrite a firm's optimization problem over technologies, for given $K > 0$ and $L(s) > 0$, for all $s \in S$ as:

$$\max_{\{D(s)\}_s} K^{\alpha}\Big[\sum_{s\in S} D(s)^{\frac{\tilde{\varepsilon}-1}{\omega\tilde{\varepsilon}}}\Big(q(s)L(s)\Big)^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}}\Big]^{\frac{(1-\alpha)\tilde{\varepsilon}}{\tilde{\varepsilon}-1}} - \sum_{s\in S} w(s)L(s) - RK, \tag{32}$$

subject to

$$\sum_{s\in S} \gamma(s)^{\omega} D(s) \leq B. \tag{33}$$

The constraint set is convex without additional restriction on parameters. Now suppose $\omega - \tilde{\varepsilon} + 1 > 0$. It implies $(\tilde{\varepsilon}-1)/\omega\tilde{\varepsilon} < 1$ because $\tilde{\varepsilon} > 1$. Under this condition, the objective function is strictly quasi-concave, so the existence and uniqueness of a global maximizer is guaranteed. Additionally, because the marginal profit of investing in $D(s)$ goes to infinity when $D(s)$ goes to zero, the solution must have $A(s) > 0$ for all $s \in S$. The symmetry of equilibrium is directly implied because all firms face the same optimization problem with unique solutions.

C.2 Long-run Elasticity of Substitution

To simplify the notation, write $\sigma = (\tilde{\varepsilon}-1)/\tilde{\varepsilon}$. Rearranging the first order condition with respect to $A(s)$ for each s gives:

$$A(s) = \gamma(s)^{\frac{-\omega}{\omega-\sigma}}\Big[q(s)L(s)\Big]^{\frac{\sigma}{\omega-\sigma}} Q^{\frac{1}{\omega-\sigma}}, \tag{34}$$

where $Q = (1-\alpha)K^{\alpha}H^{1/\tilde{\varepsilon}-\alpha}/(\lambda\omega)$ and λ is the Lagrangian multiplier. Note that (34) can also be written as:

$$\Big[\gamma(s)A(s)\Big]^{\omega} = \Big[A(s)q(s)L(s)\Big]^{\sigma} Q, \tag{35}$$

for all $s \in S$. Summing up both sides of the equation across skill groups, we have

$$B = QH^{\sigma}. \tag{36}$$

Substituting first-order condition (34) into (31), we get:

$$H = Q^{\frac{1}{\omega-\sigma}} \Big[\sum_{s \in S} \Big(\frac{q(s)}{\gamma(s)} L(s) \Big)^{\frac{\omega\sigma}{\omega-\sigma}} \Big]^{\frac{1}{\sigma}} = \Big(\frac{B}{H^{\sigma}} \Big)^{\frac{1}{\omega-\sigma}} \Big[\sum_{s \in S} \Big(\frac{q(s)}{\gamma(s)} L(s) \Big)^{\frac{\omega\sigma}{\omega-\sigma}} \Big]^{\frac{1}{\sigma}}. \tag{37}$$

Rearranging the equation, we can rewrite the aggregator H as:

$$H = B^{\frac{1}{\omega}} \Big[\sum_{s \in S} \Big(\gamma(s)^{-1} q(s) L(s) \Big)^{\frac{\omega\sigma}{\omega-\sigma}} \Big]^{\frac{\omega-\sigma}{\omega\sigma}}. \tag{38}$$

This gives the long-run elasticity of substitution:

$$\varepsilon = \frac{\omega-\sigma}{\omega-\sigma-\omega\sigma} = \frac{\omega\tilde{\varepsilon}-\tilde{\varepsilon}+1}{\omega-\tilde{\varepsilon}+1}. \tag{39}$$

Under the assumption $\omega-\tilde{\varepsilon}+1>0$, this long-run elasticity is finite and positive.

Now we derive the wage-schooling relation using the first order condition with respect to labor.

$$w(s) = \frac{\partial Y}{\partial H} H^{\frac{1}{\tilde{\varepsilon}}} \Big(A(s) q(s) \Big)^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}} L(s)^{\frac{-1}{\tilde{\varepsilon}}}. \tag{40}$$

Substituting in the first order condition regarding technology (34) gives

$$\begin{aligned} w(s) &= \lambda \omega Q^{\frac{\omega}{\omega-\sigma}} \Big(\frac{q(s)}{\gamma(s)} \Big)^{\frac{-\omega\sigma}{\omega-\sigma}} L(s)^{\frac{\omega\sigma}{\omega-\sigma}-1} \\ &= \lambda \omega Q^{\frac{\omega}{\omega-\sigma}} \Big(\frac{q(s)}{\gamma(s)} \Big)^{\frac{\varepsilon-1}{\varepsilon}} L(s)^{\frac{-1}{\varepsilon}} = \frac{\partial Y}{\partial H} \Big(\frac{H}{B^{\frac{1}{\omega}}} \Big)^{\frac{1}{\varepsilon}} B^{\frac{1}{\omega}} \Big(\frac{q(s)}{\gamma(s)} \Big)^{\frac{\varepsilon-1}{\varepsilon}} L(s)^{\frac{-1}{\varepsilon}}, \end{aligned} \tag{41}$$

which is equivalent with the first-order condition derived from the long-run aggregator.

D Immigrant Mincer Return and Cross-country Human Capital

There are two sources of efficiency associated with a schooling level—human capital accumulated from the schooling and the level of technology accessible with that schooling. To be specific, we write

$$e(s) = A(s) q(s), \tag{42}$$

where $A(s)$ denotes the efficiency from technology at level s, and $q(s) = exp(\phi_q s)$ denotes human capital at s. The gains from schooling through $q(s)$, captured by parameter ϕ_q, are typically viewed as a measure of the productivity of schooling investment. The relationship between schooling and technology, represented by $A(s)$, reflects the history of technology innovations, which may be responsive to schooling investments, as suggested by the model in Section 2.2.

In this paper, we follow Schoellman (2012) by using the Mincer returns in the United States that he estimates for U.S. immigrants as a measure of ϕ_q in the immigrants' birth countries. The intuition is that technology reflects a worker's current location, while human capital from schooling was determined by the efficiency of schooling in the country where that investment took place, that being the worker's home country.

To see this, consider the following aggregator extended from (2), where workers in the U.S. from different home countries $c \in C$ are perfect substitutes provided they have the same educational attainment.

$$H_{US} = \Big\{ \sum_{s \in S} \Big[\sum_{c \in C} e_c(s) L_c(s) \Big]^{\frac{\tilde{\varepsilon}-1}{\tilde{\varepsilon}}} \Big\}^{\frac{\tilde{\varepsilon}}{\tilde{\varepsilon}-1}}. \tag{43}$$

Immigrant workers work with U.S. technology while they accumulated human capital in their home country. The efficiency of a worker who came to the U.S. from country c is therefore:

$$e_c(s) = A_{US}(s) h_c(s) = \exp\left[(\phi_{A,US} + \phi_{q,c})s\right], \tag{44}$$

where ϕ^A denotes the projection of $A(s)$ on s. The log wage of that worker is:

$$\ln w_c(s) = \ln\left(\frac{\partial Y}{\partial H_{US}} H_{US}^{\frac{1}{\tilde{\varepsilon}}}\right) + (\phi_{A,US} + \phi_{q,c})s - \frac{1}{\tilde{\varepsilon}} \ln \tilde{H}(s), \tag{45}$$

where $\tilde{H}(s) = \sum_c e_c(s) L_c(s)$ is total efficiency-units of labor input of workers with s years of schooling. Let m_c^{US} be the Mincer return estimated in the U.S. labor market *across* immigrants from country c. From equation (45) that return is:

$$m_c^{US} = \underbrace{\left(\phi_{A,US} + \frac{1}{\varepsilon}\tilde{x}_{US}\right)}_{\zeta} + \phi_{q,c}, \tag{46}$$

where $\tilde{x}_{US}$ is the scarcity of more educated workers in the U.S. in terms of efficiency units, obtained by projecting $-\ln \tilde{H}(s)$ on s. Equation (46) shows that the cross-country variation in $\phi_{q,c}$ can be captured by the cross-home country variation that identifies m_c^{US}.