The Knowledge Trap: Human Capital and Development Reconsidered*

Benjamin F. Jones†

April 2007

Preliminary – Comments Welcome

Abstract

This paper presents a model where human capital differences, rather than technology differences, can explain several central phenomena in the world economy. The results follow from the educational choices of workers, who decide (a) how broadly and (b) how much to train, given the decisions of other workers. A "knowledge trap" occurs in economies where skilled workers favor broad but shallow knowledge. This simple idea can inform cross-country income differences, international trade patterns, poverty traps, and price and wage differences across countries in a manner broadly consistent with existing empirical evidence. Novel evidence from immigrant labor market outcomes provides further support. The model also provides insights about the brain drain and migration, and suggests an intriguing role for multinationals in assisting development. This paper shows more generally how standard human capital accounting methods can severely underestimate the role of education in development. It shows how endogenous educational decisions can replace exogenous technology differences in a range of economic reasoning.

Keywords: human capital, education, technology, TFP, relative prices, wages, cross-country income differences, international trade, multinationals, poverty traps, migration

*I wish to thank Daron Acemoglu, Pol Antras, Andrew Hertzberg, Ben Olken, Scott Stern and participants in presentations at Northwestern, Washington University and Penn State for helpful comments on initial drafts. All errors are my own.

†Kellogg School of Management and NBER. Contact: 2001 Sheridan Road, Room 628, Evanston, IL 60208. Email: bjones@kellogg.northwestern.edu.

1 Introduction

To explain several central phenomena in economics, from the wealth and poverty of nations to patterns of world trade, economists often find it necessary to invoke "technology differences". This paper suggests an alternative explanation, based in human capital. I present a model where endogenous skill endowments explain many stylized facts about the world economy – including those facts often interpreted as evidence against human capital.

The model emphasizes cross-country differences in the *quality of skilled workers*. The results follow from education decisions, which have two dimensions. One is duration - how much time to put in education - which defines whether you become a skilled worker or not. The other is content - what specific knowledge to acquire. In particular, given an investment of time, one might become a "generalist" (e.g. a generalist doctor), with modest knowledge about multiple tasks, or a "specialist" (e.g. an anesthesiologist) with deep knowledge at a particular task. Quality advantages emerge in the *collective* productivity of skilled workers, where specialists working in teams bring greater collective knowledge to bear in production.

The theory thus builds on Adam Smith's foundational observation that specialization can bring high productivity. The twist is to understand why these gains may go unrealized in the educational phase. In the model, any gains from narrowly focused training are traded off against the cost of both finding specialists with complementary skills and coordinating with them in production. This tradeoff may favor breadth over depth for three reasons. First, deep, specialized knowledge may be hard to acquire locally; for example, heart surgery may be hard to learn without guidance from existing heart surgeons. Absent expert instructors, focused training provides less advantage. Second, specialization may be worthwhile only when a sufficient mass of complementary specialists already exists. For example, learning heart surgery is less useful in the absence of anesthesiologists. These two issues suggest a variety of poverty trap where local deep knowledge is a prerequisite for individuals to willingly and successfully seek deep knowledge. Finally, coordination costs in production may be especially high.[1] For any (or all) of these reasons, a low-productivity "generalist"

[1] The idea that coordination costs of teamwork limit the gains from specialization follows Becker & Murphy (1992). More broadly, the limits to specialization considered in this paper are based on local frictions, rather than on the extent of the market as in Smith (1776).

equilibrium may persist. I call such outcomes a "knowledge trap" because the generalist equilibrium features shallower collective knowledge.

Such quality differences are limited to skilled workers. Importantly, however, the real wages of unskilled workers also rise in rich countries. This is a general equilibrium effect that follows when the output of skilled workers is relatively abundant, making the output of unskilled workers relatively scarce. This scarcity drives up unskilled wages. More precisely, when decisions to become skilled or remain unskilled are endogenous - the duration dimension of education is a choice variable - the wage structure is pinned down in equilibrium so that although quality differences are limited to skilled workers, real income gains are shared equally by skilled and unskilled workers alike.

This equilibrium effect is crucial, because it poses significant challenges to standard human capital accounting methods. The standard approach infers cross-country skill differences from within-country returns to schooling, but in this model the entire wage distribution shifts, so that within-country wage equilibria on their own say nothing about cross-country skill differences. Estimation approaches based on immigrant behavior face similar problems. The wage gains experienced by unskilled workers who immigrate from poor to rich countries need not be explained by technology, as many authors infer; in this model, the wage gains follow simply because unskilled workers are relatively scarce in the rich country. For example, one may ask why taxi drivers earn so much more in rich countries. A natural explanation is that the taxi driver's clients - skilled workers – have a much higher opportunity cost of their time and hence will pay more for the ride.

In sum, rich countries are rich because they attain deeper collective knowledge among skilled workers. The relative scarcity of low skill means that the real wages of unskilled workers also rise in rich countries, even though such workers have no more skill in rich than poor countries. One thus finds a skill-based explanation for cross-country income differences that can also get wages right.

Furthermore, the depth versus breadth tradeoff means that (a) schooling duration is insufficient to assess skill and (b) the productivity of skilled workers is interdependent. Workers are puzzle pieces, who fit together differently in different economies. This richer perspective on skill may inform phenomena such as international trade patterns (with knowledge traps, Heckscher-Ohlin can make a comeback), migrant behavior (why do skilled immigrants often take menial jobs?), and the brain drain (why don't skilled workers move to poor countries, where they are scarce?). The model further suggests pathways for escaping poverty traps, including intriguing roles for multinationals in triggering development, which may inform growth miracles in places like Hyderabad and Bangalore.

This paper is organized as follows. Section 2 introduces the core ideas. Section 3 presents a formal model, clarifying conditions for the existence of "knowledge traps" and their general equilibrium effects. Section 4 discusses several applications and relates them to existing empirical evidence in addition to new evidence about the quality of skilled workers. I show that the model provides an integrated perspective on (i) cross-country income differences, (ii) immigrant labor market outcomes, and (iii) poverty traps, as well as price phenomena, including (iv) why some goods are especially cheap in poor countries and (v) why "Mincerian" wage structures appear in all countries. The model provides additional perspectives on (vi) the role of multinationals in development and (vii) international trade patterns. Section 5 concludes.

Related Literature Many existing papers explore theoretical aspects of the division of labor (e.g. Kim 1989, Becker and Murphy 1992, Garicano 2000). Other papers explore multiple equilibria in human capital (e.g. Kremer 1993, Acemoglu 1996), and still others explore specialization in intermediate goods, i.e. at the firm level, as the source of development failures (e.g. Ciccone and Matsuyama 1996, Rodriguez-Clare 1996, Acemoglu et al. 2006). The key innovation in this paper is to imagine specialization in education as a source of multiple equilibria. More precisely, this paper imagines a two-dimensional education decision where both the breadth and duration of education are endogenous choices. There is thus a division of labor among skilled workers (based on breadth), and a division of labor between skilled and unskilled workers (based on duration).

This theoretical approach allows a reinterpretation of several empirical literatures, especially the "Macro-mincer" approach in the vast development accounting literature (surveyed in Caselli 2005), which attempts to assess the role of human capital in cross-country income differences. These empirical literatures will be discussed in detail below.

The contributions of this paper vis-a-vis existing work are primarily to (a) examine specialization failures in formal education as an explanation for persistent skill differences across countries, (b) demonstrate how standard accounting methods can severely underestimate the role of education, and (c) show a particular human capital viewpoint that provides a parsimonious interpretation of many stylized facts.

2 The Core Ideas

This section provides an introductory discussion of the core ideas in this paper. First, I introduce a "knowledge trap" to show how endogenous educational decisions can produce large cross-country differences in skill. Second, I show how standard macroeconomic accounting methods will misaccount for these skill differences. Section 3 integrates these ideas into a formal model.

2.1 A Knowledge Trap

Imagine there are two tasks, A and B, which are complementary in the production of a good. For example, the good could be heart surgery, where one task is anesthesiology and the other is the surgery itself.

Now imagine individuals must train to acquire skill, and one must decide how to use an endowment of training time. One might train as a "generalist", developing skill at both tasks. Alternatively, one might focus all their training on one task, becoming especially adept at that task. For simplicity, let training as a generalist produce a skill level 1 at both tasks, while training as a specialist produces a skill level $m > 1$ at one task and 0 at the other.

Let production be $Y = \sqrt{H^A H^B}$ when working alone and cY when pairing with another worker. This Cobb-Douglas example of a production function captures the complementarity between skills, and the term $c < 1$ represents a coordination penalty from working in a team. Output is per unit of clock-time, and the amount of skill applied to a particular task, e.g. H^A, is the summation of skill applied per unit of clock-time.

In this setting, a generalist working alone does best by dividing his time equally between tasks and earning $Y = \frac{1}{2}$. A pairing of complementary specialists optimally applies each worker to their specialty, producing $Y = mc$ for every unit of clock time, or $\frac{1}{2}mc$ per team member. The specialist organizational form is therefore more productive as long as $mc > 1$; that is, as long as coordination penalties do not outweigh the benefits of deeper expertise.

A "knowledge trap" occurs when an economy of generalists is a stable equilibrium. In a poor country, this may occur most simply because m is small. For example, becoming a skilled heart surgeon may be difficult without access to an existing skilled heart surgeon. Alternatively, coordination penalties in production may be more severe in poor countries. Hence, poor countries may feature $m'c' < 1$ while a rich country has $mc > 1$.

More subtly, an economy of generalists may persist due to thin supply of complementary specialist types. To see this, imagine you are born into a world of generalists and consider whether you would want to become a specialist instead. The best you could do as a lone specialist would be to pair with an existing generalist. In such a pairing, the specialist focuses on the task in which they have expertise, the generalist on the other, and the optimal output is $Y = \sqrt{m}c$. The generalist would have to be paid at least their outside option, $\frac{1}{2}$, to willingly join the specialist in such a team. The most income the specialist could earn is therefore $\sqrt{m}c - \frac{1}{2}$, which itself must exceed $\frac{1}{2}$ for a player to prefer training as a specialist. Hence the generalist equilibrium is stable to individual deviations if $\sqrt{m}c < 1$. We thus have a potential trap: for any coordination penalty in the range $\frac{1}{m} < c \leq \frac{1}{\sqrt{m}}$ mutual specialization is more productive and yet the generalist equilibrium is stable.[2]

I call these specialization failures a "knowledge trap" because skilled workers in the generalist equilibrium have *shallower* knowledge. This doesn't mean that they have little education. For example, the generalist doctor knows something about both anesthesiology and surgery – not to mention oncology, infectious disease, psychiatry, ophthalmology, etc.

[2] This type of knowledge trap would be resolved by mutual specialization in complementary tasks, and one may ask why this coordination problem isn't resolved naturally in the market, especially by firms. This paper assumes implicitly that educational decisions are primarily made prior to the interactions of individuals and firms, so that firms cannot coordinate major educational investments but rather make production decisions given the skill set of the labor force. This seems a reasonable characterization empirically, since skilled workers (engineers, lawyers, doctors, etc.) typically train for many years in educational institutions that are distinct from firms, before entering the workforce. In this sense, it then falls to other institutions to solve such a coordination problem. These issues will be discussed further in Section 4.

Learning something about all these different subjects may require a lot of education. But this generalist doctor will likely be far less productive than a set of specialists who work together. The specialists may have no more schooling per person, but they have much deeper knowledge about individual tasks, so that the collective body of knowledge across the specialists may be far greater. Quality differences thus follow here from the content dimension of education. To see why potentially large quality differences will not be detected by standard human capital accounting methods, we must further consider the duration dimension of education, which we turn to next.

2.2 Human Capital and Wages

A large literature has concluded that schooling variation across countries is too small to explain cross-country income differences (see Caselli 2005 for a survey). This inference is primarily drawn using the "macro-Mincer" approach, which attempts to compute human capital stocks from data on the wage-schooling relationship (e.g. Hall and Jones 1999, Bils and Klenow 2000). If workers are paid their marginal products, then the wage gain from schooling can inform how schooling influences productivity. Wage-schooling relationships are usually taken to follow the log-linear, i.e. "Mincerian", form (Mincer 1974),

$$w\left(s\right)=w(s')e^{r_m(s-s')} \tag{1}$$

where s is schooling duration, $w(s)$ is the wage, and r_m is the percentage increase in the wage for an additional year of schooling.[3]

To see how such *within-country* wage relationships can be misused in inferring *cross-country* skill differences, consider first that these wage structures emerge as a local equilibrium when labor supply is endogenous. In particular, define a worker's lifetime income as

$$y(s)=\int_s^\infty w(s)e^{-rt}dt \tag{2}$$

where individuals earn no wage income during their s years of training and face a discount rate r. If in equilibrium workers cannot deviate to other schooling decisions and be better

[3] Such log-linear wage-schooling relationships have been estimated in many countries around the world (see Psacharopolous 1994).

off, then for any two schooling levels

$$y(s) = y(s')$$

and therefore (1) follows immediately with $r_m = r$.[4] The log-linear wage structure follows through arbitrage. Individuals become skilled by investing time in education, which means giving up wages today in exchange for higher wages later. In this simple setting, the rate of return on a foregone dollar of wage income is pinned down by the expected return on investment - i.e. the discount rate.[5] Quality differences in education won't appear in the wage data, because educational duration decisions reallocate workers endogenously to ensure this equilibrium rate of return.

Now consider how one can interpret skill from wages. Imagine that there are two goods, good 1 (e.g. haircuts) produced by unskilled workers with no education and good 2 (e.g. surgery) that requires S years of training to perform. Let preferences be the same in all countries and demand for each good be downward sloping. Lastly, imagine as above that skill, h, and time, L, are the only inputs to production, so that $x_1 = h_1 L_1$ and $x_2 = h_2 L_2$. The marginal product for each good is then $w_1 = p_1 h_1$ and $w_2 = p_2 h_2$, and we have

$$h_2 = \frac{p_1}{p_2} h_1 e^{rS} \tag{3}$$

where $w_2/w_1 = e^{rS}$ follows from income arbitrage as above.

To compare skill across countries, standard accounting methods assume that unskilled workers have the same innate skill, h_1, in all economies and estimate the skill of the educated as

$$h_2 = h_1 e^{rS}$$

But this method for estimating h_2 is clearly problematic. As just shown in (3), one must also confront relative prices (p_1/p_2), which are well known to differ substantially across

[4] This arbitrage argument follows in the spirit of Mincer (1958). Integrating (2) gives $y(s) = \frac{1}{r} w(s) e^{-rs}$ so that $y(s) = y(s')$ implies $w(s) = w(s') e^{r(s-s')}$. Equivalently, (1) follows if workers choose schooling duration to maximize lifetime income. That is, with $s^* = \arg\max y(s)$ we have

$$w'(s^*) = rw(s^*)$$

which is just the log-linear wage structure expressed as a marginal condition.

[5] Here the interest rate and the return to schooling are equivalent. A richer model would introduce other aspects, such as ability differences, progressive marginal income tax rates, out-of-pocket costs for education, and finite time horizons which could drive the return to schooling above the real interest rate. See Heckman et al. (2005) for a broader characterization of lifetime income.

countries.[6] And it is easy to see how ignoring relative prices might substantially understate human capital differences. Under the innocuous assumptions that poor countries are relatively abundant in low skill and that demand is downward sloping, p_1/p_2 will be relatively small in poor countries. Hence the skill gains from education (h_2/h_1) must be adjusted upwards in rich countries relative to poor countries. Failing to account for these price differences will dampen cross-country skill differences compared to the case where we assume prices are the same everywhere.[7]

These observations suggest that skill differences might explain rather more about the world economy than a large literature has suggested. The following section presents a general equilibrium model, integrating the quality differences of knowledge traps with endogenous schooling duration decisions. Section 4 then details several applications and reconsiders established empirical evidence from the model's perspective.

3 The Knowledge Trap Economy

Imagine a world where workers are born, invest in skills, and then work, possibly in teams. They can work in one of two sectors. One sector requires only unskilled labor, and output is insensitive to the education level of the worker. Output in the other sector depends on formal education.

The key decision problem for the individual is what skills to learn. Worker type is chosen to maximize expected lifetime income. Once educated, the worker enters the labor force and produces output, which occurs efficiently conditional on the education decisions made and the ability to form appropriate teams. The educational decision is thus the key to the model.

[6] Such relative price differences are large and motivate the need for purchasing power parity (PPP) price corrections when comparing real incomes across countries.

[7] The standard accounting method assumes "efficiency wages", where the output of different skilled workers are perfect substitutes. In this case $p_1 = p_2$ (effectively, there is one good only). Under this assumption, one could estimate h_2/h_1 based purely on w_2/w_1. However, this assumption is unrealistic if we believe that worker types are less than perfect substitutes. More realistically, any number of high school students are unlikely to successfully perform angioplasty, assemble a jet engine, or write a contract consistent with the UCC. Different types of workers produce different types of goods that face downward sloping demand. Hence, skill endowments will matter in making inferences about human capital. This will be discussed further in Section 4.

3.1 Environment

There is a continuum of individuals of measure L. Individuals are born at rate $r > 0$ and die with hazard rate r, so that L is constant. Individuals are identical at birth and may either start work immediately in the unskilled sector or invest S years of time to undertake education. If they choose to educate themselves, they may develop skill at two tasks, A and B. We denote an individual's skill level $h = \{h_A, h_B\}$. An individual may choose to become a "generalist" and learn both skills, developing skill level $h = \{h, h\}$. Alternatively, one may focus on a single skill and develop deeper but narrower expertise, attaining skill level $h = \{mh, 0\}$ or $h = \{0, mh\}$ where $m > 1$.

3.1.1 Timing

For the individual, the sequence of events is:

1. The individual is born.

2. The individual makes an educational decision, becoming one of four types of workers,

   (a) Type U workers ("unskilled") undertake no education, $s^U = 0$, and have skill level $h^U = \{0, 0\}$.

   (b) Type G workers ("generalists") undertake $s^G = S$ years of education and learn both tasks, developing skill level $h^G = \{h, h\}$.

   (c) Type A workers ("A-specialists") focus $s^A = S$ years on task A, developing skill level $h^A = \{mh, 0\}$, $m > 1$.

   (d) Type B workers ("B-specialists") focus $s^B = S$ years on task B, developing skill level $h^B = \{0, mh\}$.

3. The individual enters the workforce.

   (a) Unskilled workers (type U) go to work immediately in the unskilled sector.

   (b) Skilled workers (types G, A, B) enter the skilled sector and may choose to work alone or pair with other skilled workers.

i. Unpaired skilled workers randomly meet other unpaired skilled workers with hazard rate λ.

ii. If paired and your partner dies (at rate r), then you become unpaired again.

3.1.2 Income

The expected present value of lifetime income for a worker of type k is

$$W^k = \int_{s^k}^{\infty} rV^k e^{-r\tau} d\tau \tag{4}$$

where $s^k \in \{0, S\}$ is the duration of education. Time subscripts are suppressed because we will focus on steady-state equilibria. V^k is the value of being a type k worker at the moment your education is finished, which is the expected value of being an unpaired worker of type k. This is defined by the Bellman equation,

$$rV^k = w^k + \lambda \sum_{j \in A^k} \Pr(j) \left(V^{kj} - V^k\right) \tag{5}$$

The flow value of being unpaired, rV^k, equals the wage from working alone, w^k, plus the expected marginal gain from a possible pairing. You meet other unpaired workers at rate λ, and the unpaired worker is type j with probability $\Pr(j)$. We assume a uniform chance of meeting any particular unpaired worker, so that

$$\Pr(j) = L_p^j / L_p \tag{6}$$

where L_p^j is the measure of workers of type j who are unpaired and $L_p = \sum_j L_p^j$.[8] You accept the match if $V^{kj} \geq V^k$ and reject otherwise, which defines the "acceptance set", $A^k \subset \{G, A, B\}$, the set of types that a player of type k is willing to match with. If you reject, you remain in the matching pool. If you accept, you leave the matching pool and earn V^{kj}, which is defined

$$rV^{kj} = w^{kj} - r\left(V^{kj} - V^k\right) \tag{7}$$

[8] Note that this specification guarantees that the aggregate rate at which type k people bump into type j people ($\lambda \Pr(j) L_p^k$) is the same as the rate at which type j people bump into type k people ($\lambda \Pr(k) L_p^j$). Specifically,

$$\lambda \Pr(j) L_p^k = \lambda \left(L_p^j / L_p\right) L_p^k = \lambda \left(L_p^k / L_p\right) L_p^j = \lambda \Pr(k) L_p^j$$

The flow value of being paired, rV^{kj}, is equal to the wage you receive in this pairing, w^{kj}, less the expected loss from becoming a solo worker again, which occurs when your partner dies (with probability r).

Paired workers split the value of their joint output by Nash Bargaining, dividing the joint output such that

$$w^{kj} = \arg\max_{w^{kj}} \left(V^{kj} - V^{k}\right)^{1/2} \left(V^{jk} - V^{j}\right)^{1/2} \tag{8}$$

Meanwhile, a solo worker earns the total value of his output when working alone.

3.1.3 Output

There are two output sectors. Sector 1 produces a simple good, x_1, with unskilled labor and with no advantage to skill in tasks A or B. Each worker in sector 1 produces with the technology

$$x_1 = A$$

Sector 2 produces a good where skill at tasks A and B matters. Workers in sector 2 may work alone or with a partner, with the production function

$$x_2 = Ac(n)\left(H_A^{\alpha} + H_B^{\alpha}\right)^{1/\alpha}, \; H_k = \sum_i t_i^k h_i^k \tag{9}$$

where $\sigma = \frac{1}{1-\alpha}$ is the elasticity of substitution between the two skills and we assume $\sigma \leq 1$, so that both inputs are necessary for positive production. The term $c(n) \in [0,1]$ captures the coordination penalty from working in a team of size n. Without loss of generality set $c(1) = 1$ and $c(2) = c$. The time devoted by individual i to task k is t_i^k, and members of a team split their time across tasks to produce maximum output.

3.1.4 Preferences

Utility is given by

$$U^k = \int_0^{\infty} u(C^k(t))e^{-rt}dt$$

where $u(C)$ is increasing and concave and the rate of time preference, r, is given by the hazard rate of death. Individuals have access to a competitive annuity market which pays an interest rate on loans of r.[9] The equivalence of the interest rate and the rate of time preference implies that an individual's consumption does not change across periods, by the standard Euler equation.[10] Let preferences across goods be

$$C^k(x_1, x_2) = \left(\gamma x_1^\rho + (1-\gamma)\, x_2^\rho\right)^{1/\rho} \tag{10}$$

where $\varepsilon = \frac{1}{1-\rho}$ is the elasticity of substitution between goods which we assume is positive and finite.

3.2 Equilibrium

An equilibrium is a decision by each worker that maximizes her utility given the decisions of other workers. The choice involves (a) maximizing lifetime income, and (b) maximizing utility of consumption given this lifetime income.

It will be convenient to define the equilibrium in terms of aggregate variables. Let L^k be the measure of living individuals who have chosen to be type k, and let L_q be the measure of workers actively producing the good of type q. Let X_q^S, X_q^D, and p_q respectively be the total supply, total demand, and price of good q.

Definition 1 *A full employment, steady-state equilibrium consists of W^k, C^k, L^k, L_p^k for all worker types $k \in \{U, G, A, B\}$, and L_q, X_q^S, X_q^D, p_q for each good $q \in \{1, 2\}$ such that*

1. *(Income Maximization) $W^k \geq W^j \; \forall k \in \{U, G, A, B\}$ such that $L^k > 0, \forall j \in \{U, G, A, B\}$*
2. *(Consumer optimization) $C^k(x_1, x_2) \geq C^k(x_1', x_2') \; \forall x_1, x_2, x_1', x_2'$, such that $p_1 x_1 + p_2 x_2 \leq r W^k$*
3. *(Market Clearing) $X_q^D = X_q^S \; \forall q \in \{1, 2\}$*
4. *(Steady–State) L^k is constant $\forall k \in \{U, G, A, B\}$ and L_p^k is constant $\forall k \in \{G, A, B\}$*
5. *(Full Employment) $\lambda \to \infty$*

[9] There is no capital in this model, so there is no rental rate of capital. However, there are loans, since players are born with no wealth and therefore those in school will have to borrow to consume. We imagine a zero-profit competitive annuity market where individuals hand over rights to their future lifetime income, V, upon birth in exchange for a payment, a, every period. This payment must be $a = rV$ by the zero profit condition. That is, there are rL people born per instant, meaning the annuity industry takes in rLV in income each period and there are L people alive, hence it must pay out aL. Hence, $a = rV$. Therefore, the rate of interest on loans is the same as the hazard rate of death.

[10] The Euler equation is $\frac{du'/dt}{u'} = r - r = 0$, so that $u(C)$ and hence C are constant with time.

3.3 Analysis

We analyze the equilibria in this model in two stages. First, we focus on the skilled sector. We show that two different equilibria can emerge in the organization of skilled labor, a "generalist" equilibrium and a "specialist" equilibrium. Second, we introduce the unskilled sector and demand to close the economy.

3.3.1 Organizational Equilibria in the Skilled Sector

The value of being a skilled worker of type k at the moment one's education is complete is, from (5) and (7),

$$V^k = \frac{1}{r}\frac{w^k + \frac{\lambda}{2r}\sum_{j\in A^k}\Pr(j)w^{kj}}{1+\frac{\lambda}{2r}\sum_{j\in A^k}\Pr(j)} \tag{11}$$

so that the value of being a type k worker is then a function of (a) the wage you earn if you work alone, w^k, (b) the wage you can earn in pairings you are willing to accept, w^{kj}, and (c) the rate such pairings occur, $\lambda \Pr(j)$. To solve this model, we consider the wages and pairings that can be supported in equilibrium.

Wages The equilibrium definition requires that no individual be able to deviate and earn higher income. Hence we must have $W^k = W$ for all active worker types in any equilibrium and therefore, by (4),

$$V^k = V \text{ for all } k \in \{G, A, B\}$$

That is, each type of skilled worker must have the same expected income upon finishing school. If one type did better then the others, the others would switch to become this type.

This common value, V, means that in any equilibrium individuals have the same outside option when wage bargaining. Defining x_2^{kj} is the maximum output individuals of type k and j can produce when working together, it then follows from Nash Bargaining, (8), that in any accepted pairing $V^{kj} = V^{jk}$ and

$$w^{kj} = \frac{1}{2}p_2 x_2^{kj} \tag{12}$$

so that in any equilibrium a worker team splits its joint output equally.[11] Meanwhile, if

[11] Note that $w^{kj} = p_2 x_2^{kj} - w^{jk}$. Then we can see directly from (7) that, holding V^k and V^j fixed,

skilled workers work alone, then they earn the total product, so that

$$w^k = p_2 x_2^k \tag{13}$$

where x_2^k is the maximum output an individual of type k can produce when working alone.

To proceed further, we focus on properties of the full employment setting, where $\lambda \to \infty$. Let $I(k)$ be an integer that equals the number of player types that a player of type k may successfully match with in equilibrium. To match successfully with a player of type j, a player must (a) be willing to match with that type, $j \in A^k$, and (b) be able to find such a type, $\lambda \Pr(j) > 0$.

Lemma 2 *(Full employment) With full employment,* $\lim_{\lambda\to\infty} \lambda \Pr(k) = \infty$ *if* $L^k > 0$ *and* $\lim_{\lambda\to\infty} \lambda \Pr(k) = 0$ *otherwise. Moreover,* $V^k = V^{kj}$ *in any successful match and*

$$V^k = \begin{cases} w^k/r & I(k) = 0 \\ w^{kj}/r & I(k) \geq 1 \end{cases}$$

Proof. See appendix. ■

An implication of the full employment lemma is that all skilled workers must earn the same wage in equilibrium, since common V now requires common w. To further characterize the equilibria, we must consider whether a candidate equilibrium is robust to a deviation, so we must also determine the wage one would earn if one deviates. The result is contained in the following, intuitive corollary.

Corollary 3 *(Residual Claimant) Imagine we have a full employment equilibrium. If a player deviates to become the unique member of a type k, and pairs with another player of type j, then he earns the residual value of any output from the team, net of his partner's usual wage, w. That is, the unique deviator earns a wage $w' = p_2 x_2^{kj} - w$.*

Proof. See appendix. ■

$dV^{kj}/dw^{kj} = -dV^{jk}/dw^{kj}$. It follows from the first order condition of the Nash Bargaining solution, (8), that

$$\frac{dV^{kj}/dw^{kj}}{V^{kj} - V^k} = \frac{-dV^{jk}/dw^{kj}}{V^{jk} - V^j}$$

and hence $V^{kj} - V^k = V^{jk} - V^j$. Since in equilibrium $V^j = V^k = V$, we then have $V^{kj} = V^{jk}$. Therefore, $p^2 x^{kj} - w^{kj} = w^{kj}$, so that $w^{kj} = \frac{1}{2} p_2 x_2^{kj}$.

With these results, this model possesses three noteworthy features. First, skilled workers earn their marginal products in equilibrium. Second, a player who deviates from this equilibrium would capture the full marginal product of the deviation. This property makes any resulting equilibrium in this model "robust" in the sense that deviators have every incentive to deviate in terms of their wages, since they capture the full advantage from making a different career choice. Finally, this model has a "needle in a haystack" search friction where, although search is extremely rapid ($\lambda \to \infty$), there are so many workers (a continuum) that it is still impossible to find a particular worker in finite time. This friction inhibits perfect matching and can prevent pareto optimal outcomes.

The "Knowledge Trap" Equilibria To finish characterizing the equilibria is now a matter of comparing wages in different outcomes. An equilibrium is robust if no player can individually deviate to another player type and earn a higher wage.

Consider first a world where all skilled workers are generalists. Hence $w = w^G$. If a player deviates to be a specialist, say type A, then the best he can do is pair with an existing generalist and earn $p_2 x^{AG} - w^G$. Hence, recalling that $w^G = p_2 x^G$, a world of generalists is an equilibrium iff $p_2 x^{AG} - w^G \leq w^G$, or

$$x^{AG} \leq 2x^G$$

Now consider a world of specialists, where half the measure of skilled workers are type A and the other half are type B. Here, specialists work in teams and earn a wage w^{AB}. If a player deviates to be a generalist, then he could either (a) work alone and earn w^G or (b) pair with an existing specialist and earn $p_2 x^{AG} - w^{AB}$. The latter option can never be worthwhile. In particular, since $x^{AG} < x^{AB}$, deviating to be a generalist only to pair with a specialist is never better than remaining as a specialist in the first place. We therefore only need consider the first case, where the deviating generalist works alone. Hence, this world of specialists is an equilibrium iff $w^G \leq w^{AB}$, or

$$x^{AB} \geq 2x^G$$

These existence conditions can be rewritten in terms of the model's exogenous parameters, using the production functions. The equilibria are encapsulated in the following

Proposition, which also establishes the uniqueness of these equilibria.

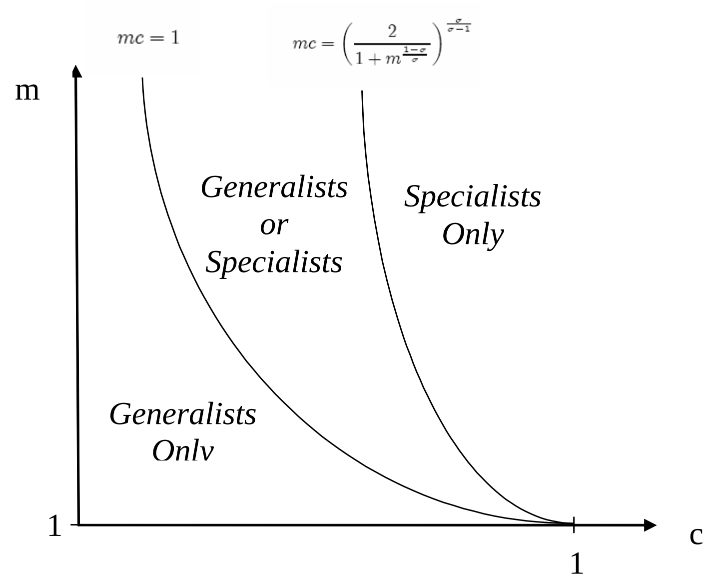

Figure 1: The Knowledge Trap

Proposition 4 *(Knowledge Trap) In a "specialist equilibrium" all skilled workers are types A or B, match with each other, and there is an equal measure of each type. In a "generalist equilibrium" all skilled workers are type G and work alone. A specialist equilibrium exists iff* $mc \geq 1$ *and a generalist equilibrium exists iff* $mc \leq \left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$*, so that for some parameter values either equilibrium is possible. A third equilibrium, where generalists and specialists coexist, occurs in the knife-edge case where* $mc = 1$*. These equilibria are summarized in Figure 1.*

Proof. See appendix. ■

Different equilibria may emerge if different economies sit in different regions of Figure 1 - i.e. if m and c differ. However, we also see that for certain regions of the parameter space, different equilibria may also emerge even if m and c are the same. This model thus can produce multiple, pareto-ranked equilibria. Moreover, the ratio of income between generalist and specialist equilibria is potentially unbounded.

Corollary 5 *(Gains from Specialization) Output in the skilled sector is mc times larger in a "specialist equilibrium" than in a "generalist equilibrium". Moreover, the range of potential* *combinations mc where both a generalist and specialist equilibria exist is unbounded from above.*

Proof. See appendix. ■

Note that with full employment, coordination costs are a necessary condition for supporting a generalist equilibrium. If there were no cost to coordination ($c = 1$) then $x^{AB} \geq 2x^G$ and a generalist equilibrium is unstable.[12]

3.3.2 The Equilibrium Economy

Given the possible organizational equilibria in the skilled sector, we know consider the influence of this organizational equilibrium on the economy at large. Denote with the subscript n the organizational equilibrium in the skilled sector, where $n = G$ defines the "generalist" outcome and $n = AB$ defines the "specialist" outcome. The equilibrium in the skilled sector will influence the endogenous outcomes in both the skilled and unskilled sectors, including labor allocations, prices, and wages.

The first result concerns wages.

Lemma 6 *(Log-linear Wages). In any full employment equilibrium*

$$w_2^n = w_1^n e^{rS} \tag{14}$$

Proof. See appendix. ■

This functional form follows from (a) exponential discounting and (b) the opportunity cost of time; it is independent of any statement about the production functions or preferences.

Given this wage relationship, we can now pin down prices. In equilibrium, workers in each sector are paid

$$\begin{aligned} w_1^n &= p_1^n A \\ w_2^n &= p_2^n 2^{\frac{1}{\sigma-1}} Ah \times \begin{cases} 1, & n = G \\ mc, & n = AB \end{cases} \end{aligned}$$

[12] If we allowed greater search frictions (finite λ), so that specialists spent periods unemployed, then this unemployment would allow the generalist equilibrium to persist even without coordination costs in production (i.e. even if $c = 1$).

Therefore, using the wage ratio, the price ratio on the supply side is determined as a function of exogenous parameters[13]

$$\frac{p_1^n}{p_2^n} = 2^{\frac{1}{\sigma-1}} h e^{-rS} \times \begin{cases} 1, & n = G \\ mc, & n = AB \end{cases} \tag{15}$$

Now consider the demand side to close the model. With CES preferences, aggregate demands are such that

$$\frac{X_1^n}{X_2^n} = \left(\frac{\gamma}{1-\gamma}\right)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{-\varepsilon}$$

Market clearing implies $p_1^n X_1^n = w_1^n L_1^n$ and $p_2^n X_2^n = w_2^n L_2^n$ so that labor allocations are also pinned down given relative prices

$$\frac{L_1^n}{L_2^n} = \left(\frac{\gamma}{1-\gamma}\right)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{1-\varepsilon} e^{rS} \tag{16}$$

where L_q^n is the measure of people actively working in sector q.[14]

Lastly, real income is also pinned down given relative prices

$$y^n = w_1^n / p^n = A \left(\gamma^{\varepsilon} + (1-\gamma)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}} \tag{17}$$

where the aggregate price level, p^n, is $p^n = \left(\gamma^{\varepsilon} (p_1^n)^{1-\varepsilon} + (1-\gamma)^{\varepsilon} (p_2^n)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$.[15]

4 Applications and Discussion

This section clarifies the implications of the model and discusses several phenomena from the model's perspective. We begin with prices - both goods' prices and wages – and build from this foundation to discuss output differences across countries, patterns of international trade, poverty traps, multinationals, and migration.

[13] The price ratio is determined entirely by the supply side because both the skilled and unskilled sectors exhibit constant returns to scale.

[14] There are also a number of students who are training in sector 2 and not yet active workers. Given the hazard rate of death r, we have $e^{rs} L_2^n$ people currently training and working in sector 2, so that total labor supply is $L = L_1^n + e^{rs} L_2^n$.

[15] Real national income (Y^n) is given by $p^n Y^n = w_1^n L_1 + w_2^n L_2$, so that real per-capita income (y^n) is

$$p^n y^n = Y^n / L = w_1^n \left(\frac{L_1^n}{L} + \frac{w_2^n}{w_1^n} \frac{L_2^n}{L}\right) = w_1^n$$

Thus average per-capita income is equivalent to the real wage in the low-skilled sector. This follows in equilibrium because workers' net present value of lifetime wage income is equivalent at birth. We can alternatively write this in terms of sector 2 wages, since $w_1^n = e^{-rs} w_2^n$.

4.1 Relative Prices

A central observation in development is that certain goods are relatively cheap in poor countries (e.g. Harrod 1933, Kravis et al. 1982). This observation motivates the need for PPP price corrections when comparing real income across countries. Explanations have been given based on exogenous cross-country differences in technology (e.g. Balassa 1964, Samuelson 1964) or factor endowments (e.g. Bhagwati 1984).

The knowledge trap model provides an endogenous explanation based on skill. Low-skilled goods (e.g. haircuts) are relatively cheap in a poor country because low skill is relatively abundant there. In particular, the price ratio is given by (15), so that

$$\frac{p_1^{AB}/p_2^{AB}}{p_1^G/p_2^G} = mc$$

and the low-skilled good is mc times cheaper in the poor country.[16]

4.2 Wages

The flexibility of prices and labor supply ensures that equilibrium wage gains from education are limited. Because people *choose* to be highly educated, excessive wage gains to the highly-educated can be arbitraged away by an increase in the supply of such workers. In the model's simple formulation, the cost of education is driven by foregone wages during the training phase. This generates the log-linear "Mincerian" wage structure and pins the skilled wage premium to the interest rate, as in (14).[17]

More generally, the model shows that with output prices and labor supply adjusting, the wage-schooling structure on its own is no longer informative of the underlying human capital in an economy, as discussed next.

[16] Relative price differences are often noted as a phenomenon of relatively cheap services (e.g. haircuts) in poor countries. One may also observe that investment goods appear relatively expensive compared to consumption goods in poor countries (Hsieh and Klenow, 2005). Testing whether skill endowments underly these cross-country price differences is an interesting empirical question left to future work.

[17] Note that this simple perspective suggests a positive correlation between interest rates and returns to schooling across countries. In fact, the literature finds both (a) higher interest rates in poor countries (e.g. Banerjee and Duflo 2004, Caselli and Feyrer, 2005) and (b) higher rates of return to schooling in poor countries (Psacharapolous 1994). See also footnote 5.

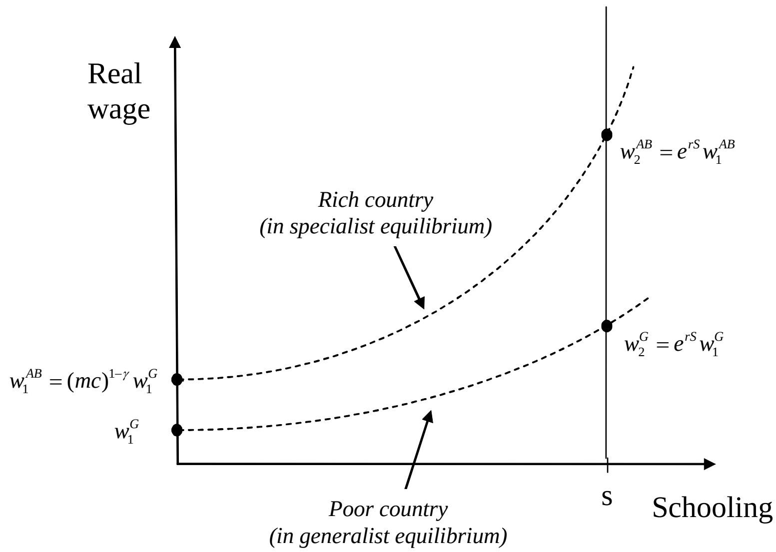

Figure 2: Equilibrium Wage-Schooling Relationships

4.3 Cross-Country Income Differences

Many authors have concluded that physical and human capital cannot explain the wealth and poverty of nations (see Caselli 2005 for a review). The residual variation in "total factor productivity" is left as a major explanation. The above model challenges these conclusions by reconsidering the role of human capital.

Typical accounting methods, following Hall and Jones (1999) and Bils & Klenow (2000), use Mincerian wage structures to count up human capital stocks assuming that different skill classes produce perfect substitutes. Under this strong assumption, the prices of any worker's output is the same and one can ignore output prices in inferring human capital. In this paper, we build instead from the viewpoint where skilled and unskilled workers are not perfectly substitutable (retail workers cannot perform heart surgery). Labor demand for different skill classes is downward sloping so that output prices adjust when the quantity or quality of a skill class changes. Countries that are very good at producing high skill will find that goods produced by low-skill workers are scarce, which drives up low-skilled wages. In fact, with relative wages pinned down by the discount rate, (14), workers will allocate themselves between skilled and unskilled careers so that the percentage wage gains for skilled and unskilled workers rise or fall in equal proportion. Wages are Mincerian in each country, but the wage-schooling relationship shifts vertically depending on the skilled equilibrium. This is shown in Figure 2 for the case of Cobb-Douglas preferences.[18]

To properly incorporate the effects of skill in a cross-country analysis, one must consider the elasticity of substitution between skilled and unskilled labor. Consensus estimates suggest an elasticity between 1 and 2 (not infinity, as the perfect substitutes approach assumes). Observing this fact, Caselli and Coleman (2006) reconsider development accounting in a setting where labor scarcity affects wages. They estimate separate productivity terms for skilled (s) and unskilled (u) workers using the production function

$$y = k^{\alpha} \left[(A_u L_u)^{\rho} + (A_s L_s)^{\rho} \right]^{\frac{1-\alpha}{\rho}}$$

which is the analogue of (10) in this paper with the addition of physical capital, k. With this estimation strategy, they find that the productivity advantage of rich countries is really limited to skilled workers. In particular, they find an enormous productivity advantage of skilled workers (A_s) in rich countries while the productivity of unskilled workers (A_u) is no higher there.[19] The "knowledge trap" model suggests exactly this effect.[20]

[18] An interesting feature of the model is that a country's average educational attainment need not even by positively associated with average income. For example, with Cobb-Douglas preferences the average schooling in a population is

$$s^n = S\frac{L_2^n}{L} = (1-\gamma)Se^{-rS}$$

a constant independent of which equilibrium is attained. For average schooling to be positively associated with income (which it is), we require the elasticity of substitution between skilled and unskilled labor to be greater than 1. Then countries with high quality skilled-labor (i.e. specialization) will see an endogenous increase in the supply of such skilled workers.

[19] In their prefered specification, Caselli and Coleman further argue that unskilled workers are actually *less* productive in rich than poor countries. An explanation for this phenomenon may be selection on ability within labor markets (see Footnote 23 below). More generally, their result for unskilled workers is sensitive to the calibration parameters and how one defines "unskilled worker". If one classifies such workers as having less than high school or less than college-level education, then unskilled workers in their calibration become mildly more productive in rich countries. What appears highly robust about their specification is that skilled workers have enormous productivity advantages in rich countries.

[20] Another important calibration is Manuelli and Sheshadri (2005), who estimate human capital by considering it as an endogenous choice variable. They find large quality differences in human capital across countries that, once accounted for, require little or no TFP differences. Their estimation suggests large advantages in the quality of education in rich countries even at entrance to primary school. This skill advantage at very low-education levels differs from the "knowledge trap" approach, which emphasizes difference that are limited to the highly skilled and differs from the Caselli and Coleman calibration, where skill differences exist only among those with more education. Manuelli and Sheshadri's imputed quality differences at all skill levels appear to follow because their model does not allow for the relative price effects that occur when skilled and unskilled workers produce different goods.

4.4 Immigrant Wages and Occupations

A second approach to assessing human capital's role in cross-country income differences is to examine what happens when workers trained in poor countries are placed in rich countries. If human capital differences were critical, authors argue, then such workers should experience significant wage penalties in the rich country's economy. Noting that immigrants from poor to rich countries earn wages broadly similar to workers in the rich country, authors have concluded that human capital plays at most a modest role in explaining productivity differences across countries (Hendricks 2002). However, this estimation approach as implemented falls into the same pitfall of standard accounting approaches, by ignoring the effect of scarce labor supply.[21]

The knowledge trap model predicts that low-skilled immigrants, who are the majority of immigrants, will enjoy (a) much higher real wages than they left behind and (b) face no wage penalty in the rich economy vis-a-vis other unskilled workers. Indeed, why would education matter for the uneducated, working as taxi drivers, retail workers, and farm hands? Wage gains follow naturally when the low-skilled immigrant moves to a place where his labor type is relatively scarce. The over-riding role of scarcity, rather than productivity, for unskilled workers is corroborated by the calibration discussed above. The potentially more informative implications of the knowledge trap model lie among skilled immigrants.

Corollary 7 *(Immigrant Workers) An unskilled worker in any country will work in the unskilled sector and earn the same wage as local unskilled workers. The skilled worker from a poor country, if he prefers to migrate, will work in the unskilled sector in the rich country and earn the unskilled wage.*

Proof. See Appendix. ■

Skilled immigrants, as generalists, are unable to find local specialists willing to team with them. Moreover, they won't work alone; the specialized equilibrium of the rich country

[21] The main estimates in Hendricks (2002) assume workers output at different skill classes are perfect substitutes thus eliminating any effect of scarcity on the wages of the unskilled. The author does consider further calibration with less then perfect substitutes, but only with elasticities of substitution between skilled and unskilled labor of approximately 5, which are far above the consensus estimates in the literature that range between 1 and 2. The calibration of Caselli and Coleman (2006) (see discussion above) uses the consensus range of elasticities, thus capturing the effects of scarce labor supply and finding no general TFP advantages in rich countries.

raises the low-skilled wage enough to make unskilled work a more enticing alternative to the immigrant generalist than using his education. Hence, for example, we will see immigrant Ph.D.'s who sometimes drive taxis.

Friedberg (2000) demonstrates that the source of education does matter to immigrant wages, but the literature does not appear to have looked explicitly at higher education. Descriptive facts can be assembled however using census data.[22] I divide individuals in the 2000 U.S. Census into three groups: (1) US born, (2) immigrants who arrive by age 17, and (3) immigrants who arrive after age 30. The idea is that those who immigrated by age 17 likely received any higher education in the United States, while those who immigrated after age 30 likely did not.

Figure 3a shows two important facts. First, controlling for age and English language ability, the location of higher education appears to matter. Among highly educated workers, those who immigrate after age 30 experience significant wage penalties, of 50% or more. Meanwhile there is no wage penalty if the immigrant arrived early enough to receive higher education in the United States. Second – and conversely – the location of high-school education does not matter. Wages do not differ by birthplace or immigration age for workers with an approximately high-school level education. Hence, the location of education matters for high skill workers but not so much for low skill workers, as the "knowledge trap" suggests.[23]

Figure 3b considers related evidence based on occupation type. To construct this graph, each occupation in the census is first categorized by the modal level of educational attainment for workers in that occupation. For example, taxi drivers typically have high school degrees, physicians typically have professional degrees, and physicists typically have Ph.D.'s. The figure shows the propensity for workers with professional or doctoral degrees to work in different occupations. We see that US born workers and early immigrants have extremely similar occupational patterns. However, late immigrants with professional or doctoral de-

[22] The data and methods are detailed further in the Appendix.

[23] Note also that immigrants with high school or less education have extremely similar wage outcomes regardless of immigration age. This further suggests that early-age immigrants are an adequate control group for late-age immigrants, highlighting that differing labor market outcomes only occur at higher education levels. Lastly, it is clear that very-low education immigrants (e.g. primary school) do significantly better then very-low educated US born workers. Such limited education is very rare among the US born and likely reflects individuals with developmental difficulties, which may explain the wage gap.

grees have a much smaller propensity to work in occupations that rely on such degrees. Instead, they tend to shift down the occupational ladder into jobs that require only college degrees and even, to a smaller extent, into occupations typically filled by those with high school or less education. This pattern is further reflected in Figure 3a, which shows that late immigrants with professional or Ph.D. degrees earn average wages no better than a locally educated college graduate.

This evidence is consistent with the "knowledge trap" model but inconsistent with a pure technology story, in which the location of education would not matter. More broadly, the evidence is consistent with the idea that human capital differences across countries exist primarily among the highly educated, as suggested independently by the calibration of Caselli & Coleman (2006) discussed above.

4.5 Poverty Traps

Unlike poverty trap models that envision aggregate demand externalities, such as big push models (e.g. Murphy et al., 1989), knowledge traps can be overcome locally, when workers achieve greater collective skill.[24] Booms are often local, whether it is city-states like Hong Kong or Singapore, or cities within countries, like Bangalore, Hyderabad, and Shenzhen, which have led growth in India and China. Yet such booms are also rare, and I consider here challenges to collective skill improvement from the perspective of the model.

4.5.1 The Quality of Higher Education

Income differences across countries may persist if countries are in different regions of Figure 1. Countries that have $mc < 1$ will have shallow knowledge and remain in poverty. This may occur if acquiring deep skills is hard in poor countries (m is small), which is reasonable if the depth acquired is limited by local access to others with deep skill - i.e. expert teachers. For example, becoming skilled at protein synthesis will be difficult without access to existing skilled protein synthesists: their lectures, advice, the ability to train in their laboratories, etc. If we write m^n, where $m^G < m^{AB}$, then countries that start in the generalist equilibrium

[24] A challenge for aggregate demand models is that many poor economies are quite open to trade or have large GDP on their own despite low per-capita GDP, so it is unclear that aggregate demand is a credible obstacle.

will remain there if $m^G c < 1$.

Escaping such a trap involves importing skill from abroad to train local students or sending students abroad and hoping they will return. Both approaches face an incentive problem however, since those with deep skills will earn higher real wages by remaining in the rich country. The model thus suggests a "brain drain" phenomenon.

Corollary 8 *(Brain Drain) Once trained as a specialist in the rich country, one will prefer to stay.*

Proof. See appendix. ■

Specialists in rich countries prefer to stay because they can work with complementary specialists there and thus earn higher wages. This result suggests that wage subsidies or other incentives may be required to attract skilled experts to the poor country and improve local training.

4.5.2 The Coordination of Higher Education

Even if poor countries can produce high-quality higher education, there is still an organizational challenge. Countries may be in the middle region of Figure 1, facing the same parameters m and c but sitting in different equilibria. Here a country cannot escape poverty without creating thick measures of specialists with complementary skills. This may be hard. Any intervention must convince initial cohorts of students to spend long years in irreversible investments as specialists, which would be irrational if complementary specialists were not expected. Hence we need a "local push".[25] Yet it is not obvious what institutions have the incentives or knowledge to coordinate such a push. A firm may have little incentive to make these investments when students can decamp to other firms.[26] Public institutions may not

[25] Some authors see such coordination failures as easily solved due to trembling hand type arguments (e.g. Acemoglu 1997). However, there are several reasons to think that small "trembles" are unlikely to undo a generalist equilibrium. First, we are considering many years of education for an individual, so that a "tremble" must be rather large. Second, while we consider two tasks for simplicity, there may be $N > 2$ tasks needed for positive output, which would then require simultaneous trembles over many specialties. Third, with greater search frictions in the market (smaller λ), trembles must occur over a large mass of workers. Fourth, in tradeable sectors, one must leap to the skill equilibrium of the rich countries to compete internationally - small skill trembles won't suffice.

[26] Contracts may help here, but labor contracts that prevent workers from departing a firm will look like slavery and are likely illegal. Labor market frictions may allow firms to do some training if frictions give the firm some monopsony power (e.g. Acemoglu and Pischke 1998). Still, it is clear that Ph.D.'s are produced in educational institutions, not in firms.

produce the right incentives either. Developing deep expertise requires time, so that the fruits of education investments may not be felt for many years, depressing the interest of public leaders (or firms), who may have short time horizons. Even if local leaders wish to intervene, it may be challenging to envision the set of skills to develop, especially if there are many required skills and deep knowledge does not exist locally. These difficulties suggest a need for "visionary" public leaders. They also suggest an intriguing role for multinationals in triggering escapes from poverty.

4.5.3 Multinationals and Poverty Traps

Intra-firm trade may play a significant role by creating production teams across national borders, and I discuss here how a multinational can play a unique role in helping countries escape poverty.

Corollary 9 *(Desirable Cheap Specialists) A firm of specialists in a rich country would hire specialists in poor countries, if they could be found.*

Proof. See appendix. ■

This follows because the skilled wage in the poor country is held down by the Mincerian wage equilibrium, making a specialist there attractive. Hence, production would shift to incorporate a skilled specialist in the poor country if such a type existed. But now we have a cross-border coordination problem. A multinational will only be able to find these specialists if they exist in sufficient measure, and no one in the poor country will want to become such a specialist unless the multinational will be able to find them.

The interesting aspect is that a multinational allows the local educational institutions to avoid producing all required specialities locally. The multinational provides the complementary worker types from abroad. For example, in Hyderabad, governor Naidu both subsidized a vast expansion in engineering education and personally convinced Bill Gates to employ these workers in Microsoft's global production chain, so that computer programmers in Hyderabad now team with other skilled specialists in advanced economies.[27] Here, the "visionary" leader need not recreate Microsoft but simply produce a sufficient quantity

[27] see, e.g., Bradsher, Keith. "A High-Tech Fix for One Corner of India", *The New York Times*, December 27, 2002, p. B1.

of one specialist type that Microsoft will hire. To the extent that a thick supply of this specialist type triggers complementary specialization locally, the local economy may escape from the trap broadly.[28]

4.6 International Trade

Finally, the knowledge trap model may provide a useful perspective on comparative advantage. The factor endowment model of trade, Heckscher-Ohlin, explains why Saudi Arabia exports oil but is famously poor at predicting trade flows based on capital and labor endowments, a failure often called the "Leontieff Paradox" (Leontieff, 1953, Maskus 1985, Bowen et al. 1987, etc).

With knowledge traps, this paradox can conceivably and naturally be resolved. Here, the rich country has a comparative advantage in the skilled good and will be a net exporter of skilled goods. Conversely, the poor country has a comparative advantage in low-skilled goods and is a net exporter of low-skilled goods.[29] Yet these comparative advantages - based in the model on specialization of training - won't appear in standard labor classifications, such as "professional" or "highly educated" worker, which can explain why attempts to save Heckscher-Ohlin through finer-grained classifications of labor types have failed (e.g. Bowen et al. 1987). With knowledge traps, rich countries are net exporters of skilled goods not simply because they have more skilled workers, but because their skilled workers have so much more skill.[30]

---

[28] With only two types of specialists, the emergence of one type in the poor country triggers the emergence of the other, and the poor country will become rich. With more than two specialist types, the emergence of one type may not inspire the local creation of the other types. Here, a multinational can continue to employ a narrow type of skilled specialists in one country without triggering a general escape from poverty. Here we will see both outsourcing and persistently "cheap engineers".

[29] In terms of the model, we can consider two small open economies who can trade both goods 1 and 2. With world prices, p_1/p_2, such that

$$\frac{p_1^G}{p_2^G} < \frac{p_1}{p_2} < \frac{p_1^{AB}}{p_2^{AB}}$$

the country in the generalist equilibrium exports the low-skilled good (1) while the country in the specialist equilibrium exports the high skilled good (2).

[30] Several studies find that augmenting factor-endowment differences with technology differences can produce more successful models empirically (Trefler 1993 and 1995, Harrigan 1997). These results, like the cross-country income literature, leave unexplained technology differences as key explanatory forces. In the view of the knowledge trap model, productivity differences are seen as the result of different, endogenous endowments of skill.

5 Conclusion

This paper offers a human-capital based explanation for several phenomena in the world economy and therefore a possible guide to core obstacles in development. The model shows how large differences in the quality of skilled labor may (a) persist across economies yet (b) not appear in the wage structure. Together, these ideas show how standard human capital accounting may severely underestimate cross-country skill differences. Building from a simple conception of human capital, the model provides an integrated perspective on cross-country income differences, poverty traps, relative wages, price differences, trade patterns, migrant behavior, and other phenomena in a way that appears broadly consistent with important facts. Future work will test the model's implications against other hypotheses on each of these dimensions.

This paper speaks directly to a long-running debate over the roles of "human capital" and "technology" in explaining income differences across countries. The model is centered on human capital, but because it directly embraces "knowledge", it also comes close to some conceptions of "technology". I close by further considering these distinctions.

In this paper, education is conceived of as the acquisition of knowledge: workers are born with no knowledge, and education is the process of embodying existing knowledge (techniques, methods, facts, theories, blueprints) into empty minds. Rich countries are rich because they load deeper knowledge into these minds than poor countries do. If we equate knowledge with "technology", then human capital and technology appear tightly related. Human capital is the embodiment of technology in the labor force, much as a microprocessor is the embodiment of technology into silicon. It is thus the emphasis on embodiment, rather than the role of "ideas", that distinguishes this paper from other approaches. By departing from the common conception of "disembodied" technology differences this paper may provide greater texture and empirical traction.

The focus on human capital does not suggest, however, that technology is not an important, distinct concept. Technology can be conceived as the finite set of discovered techniques, methods, facts, models, et cetera that limits what can be embodied in minds or machines. At the frontier of the world economy, technological progress, the expansion of this set, may still drive economic development, but even here the effects of knowledge will likely be felt – and understood – not through some disembodied process but rather through the embodiment of these ideas into the men and machines that actually produce things.

6 Appendix

Warning to reader: These proofs are preliminary and may contain typos or other errors.

Proof of Lemma (Full Employment)

Proof. The proof proceeds in three parts.

(I) In the first part of the proof I show that $\lim_{\lambda\to\infty} \lambda \Pr(k) = \infty$ if $L^k > 0$ and $\lim_{\lambda\to\infty} \lambda \Pr(k) = 0$ otherwise. I begin by defining $\Pr(k)$ in steady-state and then consider the limit of $\lambda \Pr(k)$ as $\lambda \to \infty$.

1. The probability of meeting a worker of type j is $\Pr(j) = L_p^j / L_p$. To define properties of this probability, note that unpaired workers enter and leave the matching pool by four routes. Workers enter the matching pool either because (a) they finish their studies or (b) their partner dies. Workers exit the pool either by (c) dying themselves or (d) pairing with other workers. These flows are defined as follows.

(a) There are L^k people in the population of type k. In steady state, they are born at rate rL^k and survive to their graduation with probability e^{-rs}. The rate at which new graduates enter the matching pool is therefore $rL^k e^{-rs}$.

(b) There are $L^k e^{-rs}$ type k workers in the workforce with L_p^k of these currently in the matching pool. Since workers die at rate r, the rate of reentry into the matching pool is $r\left(L^k e^{-rs} - L_p^k\right)$.

(c) Workers in the matching pool die at rate rL_p^k.

(d) Workers in the matching pool match with other unpaired workers at rate $\lambda L_p^k \sum_{j\in A^k} \Pr(j)$.

Summing up these four routes in and out of the matching pool, we have

$$\dot{L}_p^k = 2rL^k e^{-rs} - 2rL_p^k - \lambda L_p^k \sum_{j\in A^k} \Pr(j) \tag{18}$$

By the steady state assumption, $\dot{L}_p^k = 0$. With some manipulation, and using the fact that $\sum_k \Pr(k) = 1$, we can rewrite (18) as

$$\Pr(k) = \frac{d^k L^k}{\sum_j d^j L^j}$$

where $d^k \equiv \left(\frac{1}{\lambda} + \frac{1}{2r}\sum_{j\in A^k} \Pr(j)\right)^{-1}$.

2. Note that $d^k > 0$ if $\lambda > 0$. If a player type does not match, then $d^k = \lambda$. If a player does match, then d^k is no smaller than $\left(\frac{1}{\lambda} + \frac{1}{2r}\right)^{-1} > 0$. Further note that, in the limit as $\lambda \to \infty$, d^k is bounded away from zero (in fact $\lim_{\lambda\to\infty} d^k > 2r$) and d^k/λ is finite (in fact $\lim_{\lambda\to\infty} d^k/\lambda \leq 1$).

3. Since d^k is strictly positive, it therefore follows that $\Pr(k) = 0$ iff $L^k = 0$ and $\Pr(k) > 0$ iff $L^k > 0$. Moreover, writing $\lambda \Pr(k)$ as $\frac{d^k L^k}{\sum_j (d^j/\lambda) L^j}$ and recalling we have just

shown that $\lim_{\lambda\to\infty} d^k > 2r$ and $\lim_{\lambda\to\infty} d^k/\lambda \le 1$, it follows that $\lim_{\lambda\to\infty} \lambda \Pr(k) = 0$ if $L^k = 0$ and $\lim_{\lambda\to\infty} \lambda \Pr(k) = \infty$ otherwise. This gives the first part or the lemma.

(II) I next show that $V^k = V^{kj}$ in any successful match. This result follows from taking the limit of (11) as $\lambda \to \infty$, giving

$$\lim_{\lambda\to\infty} V^k = \frac{\sum_{j\in A^k} \Pr(j) V^{kj}}{\sum_{j\in A^k} \Pr(j)}$$

If k matches with only one type, then $V^k = \frac{\Pr(j)V^{kj}}{\Pr(j)} = V^{kj}$.

If k matches with two types, j and l, then $V^k = V^{kj} \frac{\Pr(j)+\Pr(l)\left(V^{kl}/V^{kj}\right)}{\Pr(j)+\Pr(l)}$. Therefore, acceptance of matches with type j is rational ($V^{kj} \ge V^k$) iff $V^{kj} \ge V^{kl}$. But similarly we can write $V^k = V^{kl} \frac{\Pr(j)\left(V^{kj}/V^{kl}\right)+\Pr(l)}{\Pr(j)+\Pr(l)}$ so that acceptance of matches with type l is rational ($V^{kl} \ge V^k$) iff $V^{kj} \le V^{kl}$. Hence, player k can only rationally accept partnerships with both types j and l iff $V^{kj} = V^{kl}$, in which case we just have $V^k = V^{kj} = V^{kl}$. We need not consider three types, because there are only three skilled types in the game and players will not match with themselves. In particular, generalists never match with generalists because $w^{GG} < w^G$. Meanwhile, a specialist never matches with his own type because $w^{AA} = w^{BB} = 0$, but $w^{kj} \ge rV^k > 0$ is a necessary condition for any match.

(III) The third part of the lemma follows directly. If you cannot successfully match, $I(k) = 0$, then it is clear from (11) that $\lim_{\lambda\to\infty} V^k = w^k/r$. If you do match, $I(k) \ge 1$, then we have just shown that $V^k = V^{kj}$. From the definition of V^{kj} in (7), it then follows directly that $V^{kj} = w^{kj}/r$. ■

Proof of Corollary (Residual Claimant)

Proof. By Nash Bargaining, $V^{kj} - V^k = V^{jk} - V^j$. Meanwhile, full employment implies $V^k = V^{kj}$ if the player k accepts matches with type j. Hence also $V^{jk} = V^j$. Now note that the outside option V^j does not change when a unique (massless) type k player appears, because we still have $\lambda \Pr(k) = 0$ and hence V^j is unchanged. Therefore, $rV^j = w$ as before, and therefore $rV^{jk} = w$. Hence $w' = x^{kj} - w^{jk} = x^{kj} - w$. ■

Proof of Proposition (Knowledge Traps)

Proof. I first show that in any equilibrium there must be (i) a positive mass of skilled workers and (ii) equal measures of type A and B specialists. I then show the conditions under which specialist and generalist organizational forms emerge.

(I) In any equilibrium, $L^A + L^B + L^G > 0$.

Assume otherwise, $L^A + L^B + L^G = 0$. Then $X_2 = 0$ and $X_1 > 0$. But this can't be an equilibrium. Under the assumption of CES preferences with finite elasticity of substitution (see (10)), $X_2/X_1 = 0$ requires $p_2/p_1 \to \infty$. But then an unskilled worker could deviate to become a skilled worker and earn an infinite wage premium, which contradicts the equilibrium condition (14).

(II) In any equilibrium, $L^A = L^B$.

Imagine an equilibrium with $L^A > 0, L^B > 0$. Consider a specialist of type A and recall that we assume $\sigma \leq 1$ so that both tasks must be performed to produce positive output. Therefore the type A worker must match, because $w^A = 0$, and she won't match with their own type, because $w^{AA} = 0$. Further, this specialist won't match with a generalist because $x^{AB} > x^{AG}$, so that $w^{AB} = x^{AB}/2 > w^{AG} = x^{AG}/2$. Therefore a type A worker will only match with a type B. By symmetric reasoning, a type B worker will only match with a type A. Therefore, the stream of payoffs to players A and B, which are equivalent in equilibrium, are $rV = \lambda \Pr(A)(V^{AB} - V) = \lambda \Pr(B)(V^{AB} - V)$. Hence, $\Pr(A) = \Pr(B)$. Hence $L^A = L^B$.

Now imagine an equilibrium with $L^A > 0, L^B = 0$. Then the best a type A worker can do is match with a generalist. If this is an equilibrium then $rV = w = x^{AG}/2$. But then a player could deviate to type B and match only with type A workers. The new wage would be $w^{'} = x^{AB} - w$. But $w' > w$ since $x^{AB} > x^{AG}$. Hence there can be no equilibrium with $L^A > 0, L^B = 0$. By symmetric reasoning, we cannot have $L^A = 0$, $L^B > 0$.

Hence, in any equilibrium, $L^A = L^B$.

(III) Now we define the equilibrium conditions for the emergence of specialist and generalist skilled workers.

1. (Specialist equilibrium) $L^A = L^B > 0$ and $L^G = 0$ is an equilibrium iff $mc \geq 1$.

Imagine that $L^A = L^B > 0$ and $L^G = 0$. If this is an equilibrium, no player would prefer to be type G. If a player deviates to type G, then instead of earning $w^{AB} = x^{AB}/2$ he earns (a) $w^G = x^G$ if he chooses to work alone or (b) $w' = x^{AG} - w^{AB}$ if he works in teams. The latter type of deviation is never worthwhile, since $x^{AG} < x^{AB} \Rightarrow w' < x^{AB}/2 = w^{AB}$. The former deviation is not worthwhile iff $w^{AB}/2 \geq w^G$, or $x^{AB} \geq 2x^G$. This holds iff $mc \geq 1$. Hence $L^A = L^B > 0$ and $L^G = 0$ is deviation proof iff $mc \geq 1$.

2. (Generalist equilibrium) $L^A = L^B = 0$ and $L^G > 0$ is an equilibrium iff $mc \leq \left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$.

Imagine that $L^A = L^B = 0$ and $L^G > 0$. If this is an equilibrium, it must be that no player would prefer to be a specialist. If a player deviates to type A, then working alone would produce $w' = 0$, so he would do best by pairing with an existing generalist and earning $w' = x^{AG} - w^G$. This deviation is not worthwhile iff $w^G \geq w'$, or $2x^G \geq x^{AG}$. This holds iff $mc \leq \left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$.

3. (Mixed specialist and generalist equilibrium) $L^A = L^B > 0$ and $L^G > 0$ is an equilibrium iff $mc = 1$.

First, as argued above As and Bs match only with each other if $L^A = L^B > 0$; hence they earn $rV = w^{AB}$. Hence, generalists must work alone, earning w^G. Hence in any such equilibrium $rV = w^{AB} = w^G$, which requires $x^{AB} = 2x^G$. Therefore, from (**??**) and (**??**),

$mc = 1$.

These results are summarized in Figure 1. ■

Proof of Corollary (Gains from Specialization)

Proof. Output per specialist is $x^{AB}/2 = mc2^{\frac{1}{\sigma-1}}h$ and output per generalist is $x^G = 2^{\frac{1}{\sigma-1}}h$, so that $(x^{AB}/2)/x^G = mc$. Hence the first part. For the second part, recall that tasks A and B are assumed to be gross complements in production ($\sigma \leq 1$), and note from the prior proposition that the upper bound on mc for generalists to exist is $\left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$, which is increasing and unbounded in m when $\sigma \leq 1$. Hence the maximum possible mc is unbounded for generalists to exist. ■

Proof of Lemma (Log-Linear Wages)

Proof. Given that individuals have the same choice set at birth and maximize income, they must be indifferent across career choices so that $W^k = W$ for all worker types, as argued above. But then, from (4), this income arbitrage means

$$\int_0^\infty w_1^n e^{-rt}dt = \int_s^\infty w_2^n e^{-rt}dt \tag{19}$$

where $w_1^n = rV^U$ is the wage paid in the unskilled sector and $w_2^n = rV$ is the wage paid in the skilled sector. Integrating (19) gives $w_n^2 = w_n^1 e^{rs}$. ■

Proof of Corollary (Immigrant Workers)

Proof. The low-skilled immigrant earns a higher real wage premium by moving to the rich country because, from (17)

$$\frac{w^{AB}/p^{AB}}{w^G/p^G} = \frac{y^{AB}}{y^G} > 1$$

Hence, from this wage consideration, the low-skilled worker in the poor country will want to migrate. This gives the first part.

Now consider the skilled immigrant.

Note first that the skilled worker in the poor country (a generalist) who migrates will never team with a specialist in the rich country. Rather, he would always prefer to work alone, since alone he earns x^G, while in a team he earns $x^{AG} - \frac{1}{2}x^{AB}$, and there are no parameter values where $x^G < x^{AG} - \frac{1}{2}x^{AB}$. (To see this, note that the condition is equivalently $1 < x^{AG}/x^G - \frac{1}{2}x^{AB}/x^G$, that $x^{AB}/x^G = mc$ and that x^{AG}/x^G is increasing in σ, attaining a maximum $x^{AG}/x^G = mc + c$ as $\sigma \to \infty$. Hence this condition requires $c > 1$, which contradicts that maintained assumption of the model that there are coordination costs in production, $c < 1$.)

Next, note that working alone as a generalist is never preferred to staying put in the poor country. The real wage of a skilled worker in the rich country must be less than mc times the real wage of skilled workers in the poor country, with strict inequality as long as there is some demand for the low-skilled good. Meanwhile, output in the skilled sector is mc times larger with specialists than with a generalist, and therefore a generalist working in the skilled sector in the rich economy will earn a wage mc times lower than the specialist workers. Hence, the real wage for the generalist skilled worker in the rich economy is lower than the real wage for the generalist in the poor economy.

Lastly, note that the generalist may still prefer to migrate and work in the unskilled sector. This occurs when the real wage gain across countries for unskilled work $\frac{w^{AB}/p_1^{AB}}{w^G/p_1^G}$ (see above) is larger than the real wage gain locally for skilled work, e^{rs}, which is more likely the greater the income differences between the countries; for example, the greater the gains from specialization, mc.

In sum, skilled workers may migrate to rich countries, but if they do they will work in the unskilled sector. ■

Proof of Corollary (Brain Drain)

Proof. The specialist who moves to the poor country will earn a wage $w' = p_2^G(x^{AG} - x^G)$. Since the poor country is in a generalist equilibrium, we must have $x^{AG} \leq 2x^G$ which implies that $w' \leq p_2^G x^G = w^G$. Hence, the skilled worker who moves from the rich to the poor country will earn a wage no greater than the skilled worker wage in the poor country. Now note that skilled workers receive a higher real wage in the rich country than the poor country because, from (14) and (17),

$$\frac{w_2^{AB}/p^{AB}}{w_2^G/p^G} = \frac{y^{AB}}{y^G} > 1$$

Hence, specialists in the rich country will prefer to stay. ■

Proof of Corollary (Desirable Cheap Specialists)

Proof. Think of the firm as a specialist in the rich country. He earns $w^{AB} = \frac{1}{2}p_2^{AB}x^{AB}$. If he can alternatively form a cross-border team with a specialist in the poor country, then he can earn at least $w = p_2^{AB}x^{AB} - p_2^{AB}x^G$, where he need provide the specialist in the poor country no more than x^G, the going rate for workers with S years of education. Hence, hiring a specialist in the poor country makes sense iff $x^{AB} - x^G \geq \frac{1}{2}x^{AB}$ or $x^{AB} \geq 2x^G$, which is just the condition for specialists to exist in the first place in the rich country. ■

Data and Analysis for Figure 3

Data on wages and occupations is taken from the 1% microsample of the 2000 United States census, which is available publicly through www.ipums.org.[31] There are 2.8 million individuals in this sample, including 320 thousand individuals who immigrated to the United States.

The wage-schooling relationships in Figure 3a are the predicted values from the following regression

$$\ln w_i = \alpha + \beta MALE + Age_{fe} + English_{fe} + Group_{fe} + Education_{fe} + Group_{fe} \times Education_{fe} + \varepsilon_i$$

where w_i is the annual wage, $MALE$ is a dummy equal to 1 for men and 0 for women, Age_{fe} are fixed effects for each individual age in years, $English_{fe}$ are fixed effects for how well the individual speaks English (the IPUMS "speakeng" variable which has 6 categories), $Education_{fe}$ are fixed effects for highest educational attainment (the IPUMS "educ99" variable, which has 17 categories) and $Group_{fe}$ are fixed effects for three different groups: (1) US born, (2) immigrants who arrive by age 17, (3) immigrants who arrive age 30 or later. Figure 3a plots predicted values from this regression, plotting the log wage against educational attainment for each of the three groups. For comparison purposes, the predicted values focus on males between the ages of 30 and 40 who speak English at least well.

To construct Figure 3b, the modal educational attainment is first determined for each of the 511 occupational classes in the data (using the IPUMS variable "occ"). Occupations are then grouped according to modal educational attainment. For example, lawyers are grouped with doctors as typically having professional degrees, and taxi drivers are grouped with security guards as typically having high school degrees. For each of the three groups defined for the $Group_{fe}$ above, Figure 3b shows the propensity of individuals with professional or doctoral degrees to work in occupations with the given modal educational attainment.

References

[1] Acemoglu, Daron. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation", *Quarterly Journal of Economics*, August 1996, 111 (3), 779-804.

[2] Acemoglu, Daron. "Training and Innovation in an Imperfect Labor Market", *Review of Economic Studies*, July 1997, 64, 445-464.

[3] Acemoglu, Daron and Jorn-Steffen Pischke. "Why do Firms Train? Theory and Evidence", *Quarterly Journal of Economics*, 1998, 113 (1), 79-120.

[4] Acemoglu, Daron, Antras, Pol and Elhanan Helpman. "Contracts and Technology Adoption", mimeo, Harvard University, 2006.

[31] Integrated Public Use Microdata Series (Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.)

[5] Balassa, Bela. "The Purchasing-Power-Parity Doctrine: A Reappraisal", *Journal of Political Economy*, December 1964, 72, 584-96.

[6] Banerjee and Duflo. "Growth Theory Through the Lens of Development Economics", *The Handbook of Economic Growth*, Aghion and Durlauf eds., Elsevier, 2005.

[7] Becker, Gary. *Human Capital*, Chicago: University of Chicago Press, 1964.

[8] Becker, Gary S. and Kevin M. Murphy. "The Division of Labor, Coordination Costs, and Knowledge", *Quarterly Journal of Economics*, November 1992, 107 (4), 1137-1160.

[9] Bhagwati, Jagdish. "Why Are Services Cheaper in Poor Countries?", *Economic Journal*, June 1984, 94, 279-285.

[10] Bils, Mark, and Peter J. Klenow. "Does Schooling Cause Growth?", *American Economic Review*, December 2000, 90 (5), 1160-1183.

[11] Bowen, Harry, Leamer, Edward and Leo Sveikauskas. "Multicountry, Multifactor Tests of the Factor Abundance Theory", *American Economic Review*, December 1987, 77 (5), 791-809.

[12] Caselli, Francesco. "Accounting for Cross-Country Income Differences", *The Handbook of Economic Growth*, Aghion and Durlauf eds., Elsevier, 2005.

[13] Caselli, Francesco and Wilbur J. Coleman. "The World Technology Frontier", *American Economic Review*, June 2006, 96 (3), 499-522.

[14] Caselli, Francesco and James Feyrer. "The Marginal Product of Capital", *Quarterly Journal of Economics*, forthcoming.

[15] Ciccone, Anotonio and Kiminori Matsuyama. "Start-Up Costs and Pecuniary Externalities As Barriers to Economic Development", *Journal of Development Economics*, 1996, 49, 33-59.

[16] Friedberg, Rachel. "You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital", *Journal of Labor Economics*, April 2000, 18 (2), 221-251.

[17] Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production", *Journal of Political Economy*, October 2000, 108 (5), 874-904.

[18] Hall, Robert and Charles Jones. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?", *Quarterly Journal of Economics*, February 1999, 114 (1), 83-116.

[19] Harrod, Roy F. *International Economics*. Cambridge Economic Handbooks, London: Cambridge University Press, 1933.

[20] Harrigan, James. "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model," *American Economic Review*, September, 1997, 87 (4), 475-494.

[21] Heckman, James J., Lochner, Lance J., and Christopher Taber. "Epxlaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogenous Agents", *Review of Economic Dynamics*, January 1998, 1 (1), 1-58.

[22] Heckman, James J., Lochner, Lance J., and Petra E. Todd. "Earnings Functions: Rates of Return and Treatment Effects: The Mincer Equation and Beyond", *IZA Discussion Paper Series* No. 1700, August 2005.

[23] Hendricks, Lutz. "How Important Is Human Capital for Development? Evidence from Immigrant Earnings", *American Economic Review*, March 2002, 92 (1), 198-219.

[24] Hsieh, Chang-Tai and Klenow, Peter. "Relative Prices, Relative Prosperity", NBER Working Paper #9701, 2003.

[25] Jones, Benjamin F. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" NBER Working Paper #11360, 2005a.

[26] Jones, Benjamin F. "Age and Great Invention", NBER Working Paper #11359, 2005b.

[27] Jones, Charles I. "On the Evolution of the World Income Distribution", *Journal of Economic Perspectives*, Summer 1997, 11 (3), 19-36.

[28] Katz, Lawrence F. and Kevin M. Murphy. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, Feburary 1992, 107 (1), 35-78.

[29] Kim, Sunwoong. "Labor Specialization and the Exent of the Market", *Journal of Political Economy*, June 1989, 97 (3), 692-705.

[30] Kremer, Michael. "The O-Ring Theory of Economic Development", *Quarterly Journal of Economics*, August 1993, 108 (3), 551-575.

[31] Leontieff, Wassily W. "Domestic Production and Foreign Trade: The American Capital Position Re-examined." *Proceedings of the American Philosophical Society*, September 1953, 97, 332-349.

[32] Manuelli, Rodolfo and Ananth Seshadri, "Human Capital and the Wealth of Nations", mimeo, University of Wisconsin, June 2005.

[33] Maskus, Keith. “A Test of the Heckscher-Ohlin-Vanek Theorem: The Leontief Commonplace”, *Journal of International Economics*, November 1985, 9, 201-212.

[34] Mincer, Jacob. "Investment in Human Capital and Personal Income Distribution", *Journal of Political Economy*, August 1958, 66 (4), 281-302.

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

[36] Murphy, Kevin P., Shleifer, Andrei, and Vishny, Robert W. "Industrialization and the Big Push", *Journal of Political Economy*, October 1989, 97 (5), 1003-1026.

[37] Pritchett, Lant. "Divergence, Big Time", *Journal of Economic Perspectives*, Summer 1997, 11 (3), 3-17.

[38] Psacharopoulos, George. "Retruns to Investment in Education: A Global Update", *World Development*, 1994, 22 (9), 1325-43.

[39] Rodriguez-Clare, Andres. "The Division of Labor and Economic Development", *Journal of Development Economics*, 1996, 49, 3-32.

[40] Kremer, Michael. "The O-Ring Theory of Economic Development", *Quarterly Journal of Economics*, August 1993, 108 (3), 551-575.

[41] Samuelson, Paul. "Theoretical Notes on Trade Problems", *Review of Economics and Statistics*, May 1964, 46, 145-54.

[42] Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*. 1776.

[43] Trefler, Daniel. "International Factor Price Differences: Leontief Was Right!", *Journal of Political Economy*, 1993, 101 (6), 961-987.

[44] Trefler, Daniel. "The Case of the Missing Trade and Other Mysteries," *American Economic Review*, December 1995, 85 (5), 1029-1046.

Figure 3a: Do Skilled Immigrants Experience Wage Penalties?
The Wage-Schooling Relationship

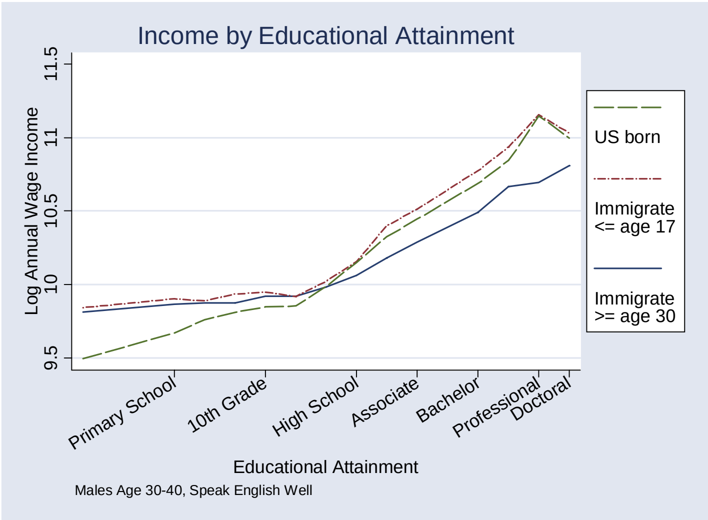

Figure 3b: Do Skilled Immigrants use their Education?
Typical Educational Level of Occupation for Workers with Professional or Doctoral Degrees