

Managerial Skills Acquisition and the Theory of Economic Development

PAUL BEAUDRY

Department of Economics, University of British Columbia and NBER

and

PATRICK FRANCOIS

Department of Economics, University of British Columbia, CIFAR and CEPR

First version received August 2006; final version accepted March 2009 (Eds.)

Why don't all countries converge rapidly to the use of most efficient or best practice technologies? Micro level studies suggest managerial skills play a key role in the adoption of modern technologies. In this paper we model the interactive process between on-the-job managerial skill acquisition and the adoption of modern technology. We use the model to illustrate why some countries develop managerial skills quickly and adopt best practice technologies, while others stay backwards. The model also explains why managers will not migrate from rich countries to poor countries, as would be needed to generate convergence. Finally we show why standard growth accounting exercises will incorrectly attribute a large proportion of managerial skills' contribution to TFP and we quantify the importance of this bias.

1. INTRODUCTION

In its purest stylized form, technology viewed as a set of instructions for bundling input combinations to produce output should be easy to transmit. Such a simple view of technology does not seem to capture well the difficulties many developing countries have had in implementing production methods that have proved successful elsewhere. A more promising embodied description emphasizes the skills required to implement technologies and the key role played by learning. Two directions that have been relatively well explored in this vein are the acquisition of skills learned through formal education and the acquisition of skills that are learnt on the job through a process of "learning by doing".

Here we emphasize an alternative embodiment process for skills acquisition which is arguably as important for the process of technology upgrading, and which has so far received less attention. We model the acquisition of (technology specific) managerial skill as resulting from on-the-job interaction with knowledgeable managers. By working in a firm and interacting with competent managers, unskilled individuals can gain managerial skills. In contrast with the literature on learning-by-doing, the skills obtained as a manager are not a substitute to the unskilled's previous skills, but instead a complement. Managerial skill is not about getting better at what one previously did, but involves learning how to organize non-managerial workers and thereby increase their productivity. Case studies of technology transfer, some of which we elaborate in the next section, are strongly suggestive of a process like this.

1 Our approach assumes that all learning activities can be internalized by the market, in the
2 sense that workers can accept lower wages to be hired in firms where they are likely to obtain
3 managerial skills. The absence of labor market imperfections causes the market outcome to
4 be characterized by an efficient hedonic equilibrium where jobs bundle both wage payments
5 and managerial learning possibilities.¹ The model generates migration incentives that are non-
6 standard, and therefore helps explain the persistence of technological backwardness. Even
7 though the model has a very neo-classical structure, the incentive for managers to migrate
8 from a more developed country (where managerial knowledge is abundant) to a less developed
9 country (where it is scarce) are shown to be weak and possibly non-existent. The wages of
10 managers can be higher in the country where they are most abundant due to the teaching value
11 component of wage payments.
12

13 The growth accounting literature has previously concluded that human capital obtained
14 on-the-job is a very minor element in explaining cross-country differences in income. In this
15 literature, the value of skills obtained on the job are proxied by the returns to experience,
16 and these do not vary sufficiently across countries to be able to explain a large part of cross
17 country variation in income (see for example, Klenow and Rodriguez-Clare 1997). We show
18 that standard methods used to account for cross-country income differences could attribute
19 most of the gain arising from differences in on-the-job acquired managerial skills to measured
20 total factor productivity; thereby suggesting that the differences are due to different availability
21 of technologies.² We provide a correction to the usual decomposition exercises and illustrate
22 how growth accounting can be modified to deliver a meaningful decomposition. For this, it is
23 necessary that the fraction of managers in the population be observable, and it is necessary to
24 use the full time-path of returns to managerial skill to properly assess the role of accumulated
25 factors versus disembodied technology. That data is not generally available, but using available
26 information, we provide a calibration of our model that demonstrates significant differences
27 in income could be driven by differences in available managerial skill across countries with
28 access to the same technology. We show that even very small differences in mortality and
29 discount rates across countries could lead to large differences in measured productivity that
30 should be attributed to managerial skill acquisition.³
31
32
33
34

35 1. The frictionless labor markets approach taken here is similar to Becker (1975). In related work, Beaudry
36 and Francois (2007), we are exploring how the results here are modified and amplified in the presence of market
37 imperfections such as in the presence of credit constraints, or thin market effects.

38 2. Growth theories that emphasize non-rivalrous technology (such as Arrow (1962a), Grossman and Helpman
39 (1991) and Aghion and Howitt(1992)) need to explain why such technologies do not diffuse rapidly to poor countries.
40 There are many proposed explanation to this puzzle. For example, one such explanation is the presence of “barriers to
41 adoption”, such as legal, institutional, socio-cultural or political factors, see Parente and Prescott (2002). Another class
42 of explanations emphasizes difficulties in absorption of new technologies. These could be due to differences in skill
43 levels, as in Acemoglu and Zilibotti (2001), which may interact with within firm contracting imperfections–Acemoglu,
44 Aghion and Zilibotti (2004). This paper offers an alternative explanation based on the process of accumulation and
45 diffusion of the managerial skills needed to use a new technology.

46 3. Since we assume all countries have access to the same technologies, the only potential source of cross-country
47 differences is individuals’ characteristics, such as differences in discount and mortality rates. While we maintain the
48 literal interpretation of these parameters throughout the paper, we believe that the analysis can also be interpreted as
49 highlighting the importance of forces that may be proxied by these parameters. For example, differences in discount
50 rates may be seen as proxying for certain frictions in financial markets, while differences in mortality rates may be
51 viewed as capturing elements related to stability of relationships.

1 This paper relates to the literature on learning-by-doing since managerial skills are assumed
2 to be acquired on the job.⁴ However, the key difference between our paper and most of the
3 learning-by-doing literature is that we do not model learning as inducing higher productivity
4 in the same task, but instead model the learning process as allowing one to obtain skills that
5 are used to organize non-managerial workers efficiently in a new technology.⁵ It also relates
6 to technology diffusion models. The closest of these is Acemoglu and Zilibotti (2001) which
7 is also concerned with explaining cross-country outcomes differences due to skill differences
8 when countries have access to the same technologies. In their framework, skill shortages lead
9 to skilled jobs being filled by the unskilled, and thus lowering productivity. They demonstrate
10 that the quantitative impact of this on measured productivity differences can be large. A key
11 difference here is that we allow for initial skill disparities to be remedied by the process of on-
12 the-job learning. Thus, in contrast to Acemoglu and Zilibotti (2001), a country's skill level is
13 an endogenous object here, and we ask whether, when allowing those skills to be augmented by
14 learning that occurs as a by-product of production, we can still generate significant cross-country
15 productivity differences. The answer we provide is yes.

16 The paper proceeds as follows. The next sub-section briefly recounts case-studies
17 exemplifying the processes that are the focus here. Section 2 sets up a model examining
18 managerial skills acquisition in a competitive economy. The steady state and transitional
19 dynamics of such a process are analyzed, with an emphasis on how the equilibrium properties
20 are affected by structural (preference) parameters. The resulting mapping between parameters
21 and outcomes offers an explanation for why some countries may adopt a technology while
22 others don't. It also explains why a small perturbation to initial condition may set off the
23 process of technological adoption. Section 3 explores the model's implications for migration.
24 In particular, this section shows why unskilled labor that moves from poor to rich countries
25 may reap high returns, and also demonstrates a good match with other elements of the empirical
26 record. Section 4 examines the consequences of this process for growth accounting. It is shown
27 that these fully internalized on-the-job skills need not be detectable in cross-country growth
28 accounting exercises, and if showing up at all, are extremely unlikely to reflect their true
29 value. Thus our model is completely consistent with observed returns to experience varying
30 little across countries. Section 5 concludes.

32 1.1. *Case studies of rapid skill acquisition*

33 As reported in Easterly (2002), in 1979 Daewoo signed a collaborative agreement with Desh
34 Garment Ltd in Bangladesh. Daewoo agreed to bring 130 Bangladeshi workers to South Korea
35 for training at a Daewoo plant in return for Desh paying commissions amounting to 8 percent
36 of future sales. At the time of the training, there were a total of 40 workers producing garments
37 in Bangladesh, and Desh's first year of operations produced \$55,050 in sales on 43,000 shirts.
38
39

40 4. Important contributions to the learning-by-doing literature include Arrow (1962b), Lucas (1993), Stokey
41 (1988) and Young (1991). A survey of the large literature on human capital and technology diffusion is provided by
42 Benhabib and Spiegel (2003). A general survey of the evidence on technology diffusion is found in Keller (2004).

43 5. In relation to the literature on technological diffusion, the mechanics of our model are closest to that of Chari
44 and Hopenhayn (1991). Their study analyzed rates of technology diffusion in a model with continuous technology
45 upgrading and vintage specificity of skills in production. In a 2-period lived overlapping generations economy where
46 the set of possible technologies expands each period, they demonstrate the existence of a stationary distribution of
47 technologies in use (with constantly increasing average productivity) and constantly increasing wages. Their analysis
48 focused explicitly on stationary, or limiting distributions, of the set of technologies in use. Instead, we characterize
the dynamics of the process by which an economy transitions from the use of a technology where managerial skills are
either not needed or already abundant, to a newer technology with higher productivity, where managerial skills are vital.

1 By 1987 the industry's output had grown to 2.3 million shirts. During the 1980s, of the 130
2 Bangladeshi workers initially trained by Daewoo, 115 of them had left Desh to set up their own
3 garment export firms. By 1995 Bangladesh was exporting \$2 billion in garments (comprising
4 54% of exports).⁶

5 Rhee and Belot (1990) documented the importance of managerial and technical skill transfer
6 in a number of successful case studies drawn from the developing world. An interesting case
7 is the meteoric rise in Indonesian plywood exports through the early 80's that started with
8 two initial firms—Korindo and Kodeco. The take-off points were joint ventures—in the case
9 of Korindo with a Korean lumber exporter, and for Kodeco with one from Taiwan. Part of
10 the knowledge transfer involved the direct relocation of managers from Korea and Taiwan, as
11 real wage increases in both countries had reduced their advantages in labor intensive plywood
12 production. The on-the-job training diffusion process in turn lead to managerial transfer beyond
13 the initial firms. In less than a decade, Indonesia moved from being an importer of plywood
14 lumber to accounting for almost 70% of the world's export share—\$2 Billion in 1976.

15 A similar process occurred in the case of Colombian flower exporting. A single company,
16 Floramerica, undertook a feasibility study on exporting flowers to the US in 1969. Prior to
17 that time there were no commercial flower enterprises in the country. Within a year it had
18 150 employees, and high rates of growth continued. The \$100,000 initial investment grew into
19 a company with \$50 million in annual sales by 1986. In addition to Floramerica's dramatic
20 growth, the flower producing and marketing know-how diffused to others in the Colombian
21 economy. In less than two decades, flowers became Colombia's fifth largest export, and in the
22 early 90's Colombia was second only to the Netherlands in supplying cut-flowers to the world
23 market. At the period of study, flower production for export directly employed 70,000 people
24 in Colombia, Rhee and Belot (1990, p31).⁷

25 The means by which the relevant skills are acquired in these examples—on the job—is
26 consistent with evidence from the human resources literature. The know-how that is most
27 important for management seems difficult to codify and learn through formal institutions.
28 Most managerial learning takes place informally in the workplace—see Enos, Kehrhahn, and
29 Bell (2003) for a recent study and survey of the managerial studies evidence. For example,
30 McCall, Lombardo, and Morrison (1988) found that 30 of 35 managerial job skills were learned
31 informally, with relatively little gained from formal training procedures. The development
32 of proficiency in a skill is a process of tacit learning which seems to depend critically on
33 knowledge gained through experience, and from observing actual workplace leaders at work.⁸
34 These features are the key elements to the model that we now develop.

35 2. THE MODEL

36 *Households*

37 Consider an economy where the population (or labor force) is of constant size and
38 normalized to 1. At each instant a new cohort of size δ enters and a measure δ of workers dies;

39
40
41
42 6. The first detailed case study of Desh was undertaken by Rhee and Belot (1990). Examples of technology
43 diffusion via skilled worker and managerial mobility are commonplace in modern economies. Two examples from
44 advanced sectors are Saxenian (1994), who documents diffusion in Silicon Valley's IT clusters, and Feldman's (2001)
45 study of firm formation in biotechnology and communications industries in Capitol region clusters. Lall (1999)
46 documents the dramatic increases in the use of managers in the fast-growing East-Asian economies.

47 7. A more recent study by Lester and Piore (2004) estimates that upto 30% of India's IT workforce rotates
48 through American firms to upgrade their tacit knowledge.

8. Kerr (2008) estimates the role of ethnic ties in helping this type of skill transfer, and hence in facilitating
technology diffusion for the case of ethnic scientific communities working in the US.

1 with the property that $\int_{-\infty}^t \delta e^{-\delta(t-s)} ds = 1$. Each individual is risk neutral and inelastically
 2 supplies one unit of labor per instant. Let $\rho > 0$ be the instantaneous discount rate. All newly
 3 entering workers do not have managerial skills, and are thus called unskilled. We will often
 4 refer to managers as skilled workers. Discounted expected lifetime utility for an individual is
 5 $\int_0^{\infty} c(t) e^{-(\delta+\rho)t} dt$, where $c(t)$ denotes consumption.⁹

6 *Production*

7 Production comprises two types of constant returns to scale activity; denoted traditional
 8 and modern. In the modern production process there are two types of positions: management
 9 positions and unskilled worker positions. The measure of skilled managers is denoted S , while
 10 the measure of unskilled workers is denoted U . In the traditional technology, there is only one
 11 type of position which we call laborers and denote by L . In order to be a manager, an individual
 12 must possess managerial skill. In contrast, anyone can be hired as an unskilled worker in the
 13 modern technology or as a laborer in the traditional technology.

14 Total output produced using S skilled managers, U unskilled workers, and L laborers
 15 is given by $Y = F(S, U) + \bar{w}L$, where \bar{w} is labor's marginal product in the traditional
 16 technology.¹⁰ We impose the following Inada type condition on production:

$$17 \quad \text{for } U > 0, \quad \lim_{S \rightarrow 0} F_1(S, U) = \infty \text{ and } F_2(0, U) = 0,$$

$$18 \quad \text{for } S > 0, \quad \lim_{U \rightarrow 0} F_2(S, U) = \infty \text{ and } F_1(S, 0) = 0.$$

$$21 \quad \text{by continuity then } \exists S' : F_1(S', 1 - S') = F_2(S', 1 - S'). \quad (1)$$

23 A natural interpretation of this last condition is that production necessitates some individuals
 24 being placed in positions requiring organizational skills, i.e., management, but when all such
 25 positions are occupied, the skilled are more usefully allocated to positions that could easily be
 26 filled by the unskilled.¹¹

27 If the modern technology is not managerial skill constrained, the marginal unskilled
 28 individual is more productive working with it than in the traditional technology, i.e.:

$$30 \quad F_1(S', 1 - S') = F_2(S', 1 - S') > \bar{w}. \quad (2)$$

32 *On-the-job acquisition of managerial skills*

33 Unskilled individuals working in the modern technology acquire managerial skills randomly
 34 following a Poisson process. The instantaneous rate of learning, denoted $\Omega(S, U)$, depends on
 35 both the number of managers and the number of workers in a firm.¹² In particular, we assume
 36 that $\Omega_1 \geq 0$, $\Omega_2 \leq 0$ and Ω is homogeneous of degree zero. The homogeneity of degree

38 9. The two parameters, ρ and δ , enter identically here but will be seen to play markedly different roles in
 39 governing the behavior of the system in what follows. The difference arises from the additional impact of δ on the
 40 persistence of skills and the inflow of unskilled.

41 10. Note that capital plays no role in the analysis and that this is without loss of generality.

42 11. In the Cobb-Douglas case, $S^\alpha U^{1-\alpha}$, this arises when $S > S'$ where $S' = \alpha^\alpha (1 - \alpha)^{1-\alpha}$. In the traditional
 43 technology, we do not make a distinction between managers and unskilled workers. This scenario can be interpreted
 44 as a situation where the traditional technology has already reached its level of S' , so that the distinction between
 45 managers and unskilled in this technology is immaterial.

46 12. An alternative assumption regarding managerial skills acquisition is that an unskilled individual's learning
 47 rate is increasing in the number of "active" managers in the firm, not the number of managers, S . This distinction
 48 arises because it will be seen that, in some configurations, firms do not employ all potential managers in managerial
 49 positions. All of the paper's results are preserved under this alternative formulation, and we proceed with the simpler
 50 formulation stated in the text.

1 zero implies that the speed of learning in a firm depends only on the ratio of managers to
 2 workers. We will also employ the notation $\tilde{\Omega}(S, U)$ to denote the function $\Omega(S, U)U$ which
 3 is the instantaneous skilled output of the learning production process. Clearly $\tilde{\Omega}(S, U)$ is
 4 homogenous of degree one and we further impose the restriction that it is a concave function.
 5 These assumptions imply that $\frac{\partial \tilde{\Omega}}{\partial U} = \Omega + U\Omega_2 > 0$, or in words, that increasing the number
 6 of unskilled individuals working with the modern technology increases the diffusion of the
 7 managerial skill.

8 In order to allow the process of adoption of the modern technology to begin with $S = 0$,
 9 we assume that managerial skills can be learned by working with the new technology even
 10 without the presence of other managers, that is, we assume that $\Omega(0, x) > 0$ for $x > 0$.
 11 A skilled individual remains skilled throughout their life - i.e., individual skills do not
 12 depreciate.

14 2.1. The decentralized equilibrium

16 Managerial skills acquisition is a by-product of production but need not imply externalities,
 17 since skills acquisition is excludable. Individuals must work in the modern technology to
 18 acquire managerial skills, and do so more quickly the more managerial input they receive.
 19 This teaching aspect allows firms with a high ratio of managers to offer a low wage to
 20 unskilled workers since they can learn managerial skills more quickly in such a firm. Thus
 21 profit maximizing firms make employment decisions taking into account that their ratio of
 22 managers to unskilled workers affects the unskilled wages they can offer. An employment
 23 offer to an unskilled individual is actually a bundle that specifies a wage, w_t^u and a (firm
 24 specific) speed of learning Ω . Unskilled workers accordingly choose the bundles that maximize
 25 utility. We characterize the hedonic competitive equilibrium (see Rosen (1974)). This is a
 26 situation where both firms and workers take the market tradeoff between w^u and Ω as given
 27 when making their decisions, with the equilibrium requirement that demand equals supply
 28 at each chosen pair of w^u and Ω . Here, additionally, hedonic equilibrium is dynamic as
 29 there is a sequence of tradeoffs between w^u and Ω , one at each moment in time. It will
 30 be useful to denote the market tradeoff between w^u and Ω at time t by the function $g_t(\cdot)$,
 31 with

$$32 \quad w_t^u = g_t(\Omega), \quad g_t'(\cdot) \leq 0. \quad (3)$$

35 A competitive hedonic equilibrium specifies a sequence of managerial wages, w_t^s and
 36 a sequence of unskilled-wage-learning tradeoffs $w_t^u = g_t(\Omega)$, such that given these trade-
 37 offs, firms choose employment to maximize profits, households choose their allocation of
 38 labor to maximize utility, and the markets for managers, unskilled workers and laborers
 39 clear.

41 **2.1.1. The Firm's Problem in a Hedonic Equilibrium.** Firms are competitive, and can
 42 costlessly enter and exit instantaneously. They thus simply maximize static profits by choosing
 43 the number of skilled workers to place in managerial positions, the number of unskilled workers,
 44 and the wage to pay unskilled workers w_t^u . They take as given the market wage of skilled
 45 workers, w_t^s , and the market tradeoff $g_t(\cdot)$ between w_t^u and Ω . A firm hiring S_t skilled workers
 46 places a fraction $\gamma_t \leq 1$ into managerial positions, and the remainder, $(1 - \gamma_t) S_t$, work with the
 47 unskilled, U_t . Firms solve $\max_{\gamma, S, U, w^u} (F(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - w_t^s S_t - w_t^u U_t)$ subject to
 48 $w_t^u = g_t(\Omega(S_t, U_t))$, and $0 \leq \gamma_t \leq 1$, i.e., firms recognize how their choice of S_t and U_t

affects the speed of learning $\Omega(\cdot)$. This yields first order conditions:

$$w_t^s = F_1(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - g'_t(\Omega) \Omega_1(S_t, U_t) U_t \quad (4)$$

$$w_t^u = F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - g'_t(\Omega) \Omega_2(S_t, U_t) U_t \quad (5)$$

$$0 = [F_1(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t)] (1 - \gamma_t) \quad (6)$$

$$F_1(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) \geq F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) \quad (7)$$

Note that since $\Omega_1 > 0$ the wages of managers reflect their teaching value, whereas $\Omega_2 < 0$ reflects the congestion effect on learning of the presence of too many workers.

2.1.2. The Household's Problem in a Hedonic Equilibrium. Individuals with managerial skill at time t , denoted S_t , receive the market wage for skilled, w_t^s , given in equation (4). Though it is possible for these individuals to remain in the old technology receiving \bar{w} , since this is never part of an optimal program for the skilled, to save on notation, we do not denote this possibility.

An unskilled worker can work in the new technology probabilistically acquiring managerial skills and receiving wage w_t^u or remain in the traditional technology receiving \bar{w} . In the old technology he remains unskilled. In the new, the instantaneous rate of skill acquisition is given by $\Omega(S_t, U_t)$, and firms offer pairs (w_t^u, Ω) , that can potentially vary by firm.¹³ Since the unskilled may be indifferent between remaining in the old technology and working in the new, we allow them to follow mixed strategies. Let β_t ($0 \leq \beta \leq 1$) be the probability that an unskilled individual chooses to work in the new technology. Using Bellman's Principle of optimality, with V_t^U representing the expected discounted utility of being currently unskilled, and V_t^S representing the expected discounted utility associated with being skilled, we have that $(\delta + \rho)V_t^U = \max_{\beta_t, \Omega} (\beta_t w_t^u + (1 - \beta_t)\bar{w} + \beta_t \Omega (V_t^S - V_t^U) + V_t^U)$ subject to $w_t^u = g_t(\Omega)$, and $0 \leq \beta_t \leq 1$.

An unskilled individual recognizes that working with the new technology yields an Ω chance of attaining skills and transitioning to the skilled state. He also recognizes the trade-off offered by the market in terms of wages in the modern sector and the chance of learning to manage a modern technology. The skilled individual does not care about learning probabilities directly, but will internalize these according to the hedonic relationship described by g . The unskilled also care directly about their probability of acquiring skills, Ω .

The associated FOCs associated with the household's problem are:

$$(w_t^u - \bar{w} + \Omega(S_t, U_t)(V_t^S - V_t^U)) \begin{cases} = 0 & \text{if } 0 < \beta_t < 1 \\ \leq 0 & \text{if } \beta_t = 0 \\ \geq 0 & \text{if } \beta_t = 1 \end{cases}$$

$$g'(\Omega(S, U)) = -(V_t^S - V_t^U). \quad (8)$$

In the above expression of the first order conditions, we have replaced the choice variable Ω with its equilibrium value given by $\Omega(S_t, U_t)$. These conditions imply that when $0 < \beta < 1$, V_t^U can be rewritten as $(\delta + \rho)V_t^U = w_t^u + \Omega(S_t, U_t)(V_t^S - V_t^U) + V_t^U$. Again using Bellman's principle, V_t^S can be written as $(\delta + \rho)V_t^S = w_t^s + V_t^S$. Taking the difference

13. To save on notation, we do not explicitly denote firm level Ω_s , as all will choose the same Ω , w^u pairs in equilibrium.

1 between these two last equations, and denoting the shadow value of being skilled by λ_t , (i.e.,
2 $\lambda_t = V_t^S - V_t^U$), we get $(\delta + \rho)\lambda_t = w_t^s - w_t^u - \Omega(S_t, U_t)\lambda_t + \dot{\lambda}_t$.

3 More generally, for β either in the interior or at the boundaries, λ_t will need to satisfy

$$4 \quad (\delta + \rho)\lambda_t = w_t^s - \beta_t w_t^u + (1 - \beta_t)\bar{w} - \beta\Omega(S_t, U_t)\lambda_t + \dot{\lambda}_t \quad (9)$$

6 where β_t now represents the optimal values of β . In addition to (9), λ will need to satisfy the
7 transversality condition:

$$9 \quad \lim_{t \rightarrow \infty} \exp^{-\rho t} \lambda_t = 0 \quad (10)$$

11 In the case where there is an interior solution ($0 < \beta < 1$), the optimality condition (2.1.2)
12 has a simple interpretation. It states that the current loss between working as unskilled in the
13 new technology as opposed to working in the traditional technology must be compensated by
14 the expected gain associated with potentially learning skills, since in this case the gain to such
15 skills is λ_t and the probability of acquiring them is $\Omega(S_t, U_t)$. Equation (9) also has a simple
16 interpretation, especially in a steady state situation with $\dot{\lambda} = 0$. For example, when $\beta = 1$ in the
17 steady state, $\lambda = \frac{w_t^s - w_t^u}{\delta + \rho + \Omega}$, which implies that the value of being skilled is simply the actualized
18 difference between receiving w_t^s and w_t^u , where the discount factor incorporates the fact that
19 an unskilled person may become skilled.

21 2.2. Learning and accumulation dynamics

23 The dynamic equation governing evolution of skills depends on the fraction of unskilled
24 individuals supplying their work time to firms using the new technology $\beta_t(1 - S_t)$ and the
25 rate at which one learns, $\Omega(S_t, \beta(1 - S_t))$. Hence, the dynamics of S is given by

$$27 \quad \dot{S} = \Omega(S_t, \beta(1 - S_t))\beta(1 - S_t) - \delta S_t \quad (11)$$

29 2.3. Statement of equilibrium

31 An Equilibrium is composed of a sequence of skilled wages, unskilled wages and of a (differ-
32 entiable) learning-wage tradeoff $\{w_t^s, w_t^u, g_t(\cdot)\}$ and a sequence of allocations $\{\beta_t, \gamma_t, S_t, \Omega_t\}$
33 such that:

- 34 1) Given $\{w_t^s, w_t^u, g_t(\cdot)\}$, Ω_t and β_t solve the households decision problem with respect to
35 occupational choice.
- 36 2) Given $\{w_t^s, w_t^u, g_t(\cdot)\}$, $\gamma_t, S_t, U_t, \Omega_t$ solve the firms' problems.
- 37 3) S_t obeys equation (11).
- 38 4) The market for workers clears: $\beta_t(1 - S_t) = U_t$.
- 39 5) $w_t^u = g_t(\Omega_t)$ and $\Omega_t = \Omega(S_t, U_t)$

41 2.4. Analysis

43 Substituting for wages from (4) and (5), into the first order conditions given by (2.1.2) and
44 substituting out for g by using (8) yields

$$47 \quad \bar{w} - F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - \lambda_t \Omega_2(S_t, U_t) U_t \begin{cases} \geq \lambda_t \Omega(S_t, U_t) & \text{if } \beta_t = 0 \\ = \lambda_t \Omega(S_t, U_t) & \text{if } \beta_t \in (0, 1) \\ \leq \lambda_t \Omega(S_t, U_t) & \text{if } \beta_t = 1 \end{cases}$$

1 and $\dot{\lambda}_t - (\delta + \rho)\lambda_t = \beta_t (F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) + \lambda_t \Omega_2(S_t, U_t) U_t) + (1 - \beta_t)\bar{w} - F_1$
 2 $(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - \lambda_t \Omega_1(S_t, U_t) U_t + \lambda_t \beta_t \Omega(S_t, U_t)$. If S_t is such that $F_1(S_t, 1 - S_t)$
 3 $\geq F_2(S_t, 1 - S_t)$ managerial skills are valuable and all managers are placed in manage-
 4 rial positions, $\gamma = 1$. Unskilled individuals may then decide to accumulate skills, $\beta \geq 0$. If
 5 $F_1(S_t, 1 - S_t) < F_2(S_t, 1 - S_t)$ then the unskilled in the new technology are more productive
 6 than in the old (via condition 2) and firms optimize their allocation by setting $\gamma < 1$, so that
 7 $F_1(\gamma S_t, (1 - S_t) + (1 - \gamma) S_t) = F_2(\gamma S_t, (1 - S_t) + (1 - \gamma) S_t)$.

8 In summary, the problem of characterizing the market equilibrium can be reduced to
 9 characterizing the following system in β_t, γ_t, S_t and λ_t , where we either have scarce and
 10 valuable skills, $\gamma = 1$ or skills are abundant and of zero value, $\gamma \leq 1$.

11 EITHER

$$12 \quad F_1(S_t, 1 - S_t) - F_2(S_t, 1 - S_t) \geq 0$$

13 AND

$$14 \quad \gamma = 1$$

$$15 \quad \bar{w} - F_2(S_t, \beta_t(1 - S_t)) + \lambda_t \beta_t (1 - S_t) \Omega_2(S_t, \beta_t(1 - S_t))$$

$$16 \quad \begin{cases} \geq \lambda_t \Omega(S_t, 0) & \text{if } \beta_t = 0 \\ = \lambda_t \Omega(S_t, \beta_t(1 - S_t)) & \text{if } \beta_t \in (0, 1) \\ \leq \lambda_t \Omega(S_t, 1 - S_t) & \text{if } \beta_t = 1 \end{cases} \quad (12)$$

$$17 \quad \dot{\lambda}_t - (\delta + \rho)\lambda_t = \beta_t F_2(S_t, \beta_t(1 - S_t)) + \beta_t \lambda_t \Omega_2(S_t, \beta_t(1 - S_t)) \beta_t(1 - S_t)$$

$$18 \quad + (1 - \beta_t)\bar{w} - F_1(S_t, \beta_t(1 - S_t))$$

$$19 \quad - \lambda_t \Omega_1(S_t, \beta_t(1 - S_t)) \beta_t(1 - S_t) + \lambda_t \beta_t \Omega(S_t, \beta_t(1 - S_t))$$

20 OR

$$21 \quad F_1(\gamma S_t, (1 - S_t) + (1 - \gamma) S_t) = F_2(\gamma S_t, (1 - S_t) + (1 - \gamma) S_t)$$

22 AND

$$23 \quad \gamma < 1, \lambda_t = 0, \beta_t = 1$$

24 with the addition that the transversality condition (10), and the aggregate accumulation
 25 equation (11), must hold.¹⁴

26 The above system of equations in β_t, γ_t, S_t and λ_t is complicated since it involves different
 27 regimes. However, most of its properties can be illustrated graphically using a phase diagram
 28 in the $S - \lambda$ space once it is recognized that β_t and γ_t can be solved as functions of S_t and λ_t
 29 (i.e., β_t and γ_t are determined by static conditions, reducing the dimension of the problem to
 30 a dynamic system in only two variables). In order to understand the properties of this system,
 31 it is helpful to begin by distinguishing the regions in the $S - \lambda$ space where $\beta = 1$ and $\gamma = 1$
 32 from those where these two variables are strictly smaller than 1.

33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 14. It can be shown that the set of equilibrium requirements can be derived as the first order conditions of a
 social planner's problem which maximizes the discounted utility of consumption subject to the resource constraint
 and the dynamics of S . This equivalence can be used to prove that the equilibrium is Pareto optimal and is available
 from the authors upon request.

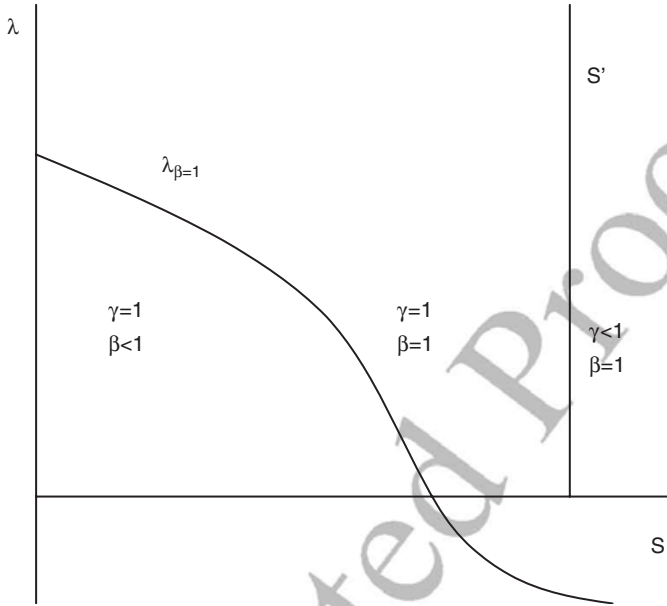


FIGURE 1

2.5. The γ and β regions

Figure 1 depicts the $S - \lambda$ space being cut by two lines. The vertical line at S' is implicitly defined by the expression $F_2(S, (1 - S)) = F_1(S, (1 - S))$ (equation (2)). To the left of this line $\gamma = 1$, that is, to the left, all individuals with the skills to be managers are employed by firms as managers since their marginal product is higher there. To the right $\gamma < 1$, since managerial abundance implies it is optimal to have some potential managers employed as non-managers.¹⁵

The second line in Figure 1 depicts the set of points that delimit the region where $\beta = 1$ versus when $\beta < 1$. This relationship is given by $\lambda = \frac{\bar{w} - F_2(S, (1 - S))}{\Omega + \Omega_2(1 - S)}$ and is derived from condition (12). We will denote this relationship as $\lambda_{\beta=1}$ from now on. The region to the right of this line represents the region where it is optimal to have all unskilled workers working in the modern technology. In contrast, the region to the left of this line is where it is optimal to leave some unskilled workers working with the traditional technology. In the appendix it is proved that this schedule is downward sloping. Intuitively, this line yields the minimal level of managerial skills required for an economy to be willing to fully devote itself to the modern technology. It slopes downwards because, for higher S , the marginal product, and hence wage of the unskilled in the new technology rises, so that a lower shadow value of skills, λ , is sufficient to induce the unskilled into the modern sector.

The $\lambda_{\beta=1}$ line plays a particularly important role since the dynamic properties of the system, as reflected in the $\dot{S} = 0$ and $\dot{\lambda} = 0$ lines differ depending on which side of this line they lie. The next step is therefore to characterize the dynamics of S and λ globally. The details of this characterization are found in the appendix with the salient results regarding the shape of the critical $\dot{\lambda} = 0$ and $\dot{S} = 0$ schedules summarized below:

15. Note that potential managers never choose the traditional technology since the modern technology is superior.

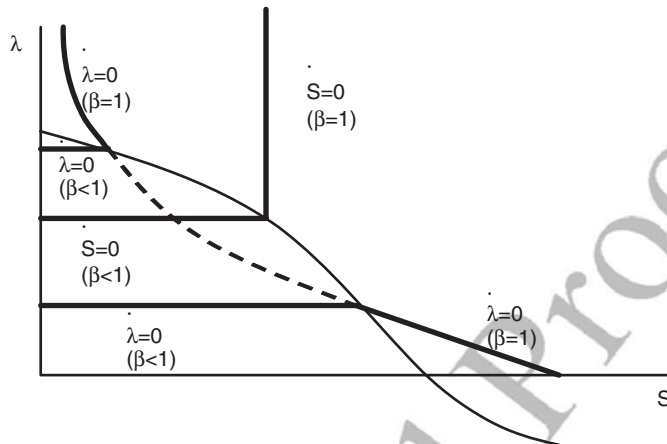


FIGURE 2

Lemma 1. *In $S - \lambda$ space, as depicted in figure 1: (1) For λ values such that $\beta = 1$: The $\dot{\lambda} = 0$ line is negatively sloped and the $\dot{S} = 0$ line is vertical. (2) For λ values such that $\beta < 1$: The $\dot{\lambda} = 0$ line is horizontal and the $\dot{S} = 0$ line is horizontal.*

Figure 2 summarizes the lemma's implications for the shapes of such schedules. Both schedules are depicted by the heavier lines. Both are horizontal for $\beta < 1$, while in the $\beta = 1$ range the $\dot{\lambda} = 0$ line slopes downward and the $\dot{S} = 0$ line is vertical.

The $\dot{\lambda} = 0$ condition depicts the set of points for which there is balance between the forces causing directional change in λ , the shadow value of skills. In the $\beta = 1$ region, a higher value of S , and hence a lower value for the differential $F_1(S, (1 - S)) - F_2(S, (1 - S))$, must be offset by a lower value of λ . Intuitively, as skills become less valuable instantaneously, the shadow value of being skilled, λ , must fall. In the $\beta < 1$ region, this relationship is more subtle, because as S changes, there is also an endogenous change in β , as reflected in condition (12). Intuitively, as more managers are available in the new technology, both the effectiveness of learning and the productivity of the unskilled rises, ceteris paribus. However, since the outside option has remained constant (i.e., working in the old technology) this would induce proportionately more unskilled to enter the new technology (β rises). The $\dot{\lambda} = 0$ locus is horizontal in this region because increases in the numbers of skilled managers are being perfectly offset by increases in the unskilled that are willing to work with them there, due to linear homogeneity. This remains so until there are no more unskilled available, $\beta = 1$, after which the relationship is again downward sloping.

The $\dot{S} = 0$ schedule is also relatively simple to understand in the $\beta = 1$ region. In that region, there can be balance in the forces affecting S if and only if attrition in managers (δS) just offsets growth in managerial skills ($\Omega(S, 1 - S)(1 - S)$) as given by $\dot{S} = 0$ in equation (11) with $\beta = 1$. This is a relationship that is independent of skills' shadow value, λ , and hence the vertical slope of that schedule in this region. For $\beta < 1$ this relationship is again complicated by changes in β induced by varying S . For balance between the inflow, $\Omega(S_t, \beta(1 - S_t))\beta(1 - S_t)$, and outflow of managers in the new technology, δS_t , linear homogeneity implies that the manager/worker ratio cannot change with S . Since this ratio also determines the effectiveness of learning the new technology, the effectiveness of learning

cannot change with S , and since the outside option to learning (the wage in the old technology) is constant, the shadow value of skills, λ , must also be constant with S .

2.6. Steady states and transitional dynamics

Steady states occur at intersections of the $\dot{S} = 0$ and $\dot{\lambda} = 0$ schedules. Although $\dot{S} = 0$ and the $\dot{\lambda} = 0$ locus each have two possible configurations depending on the regions in which they lie, as described above, it is possible to show that only three qualitatively distinct steady state configurations can arise:

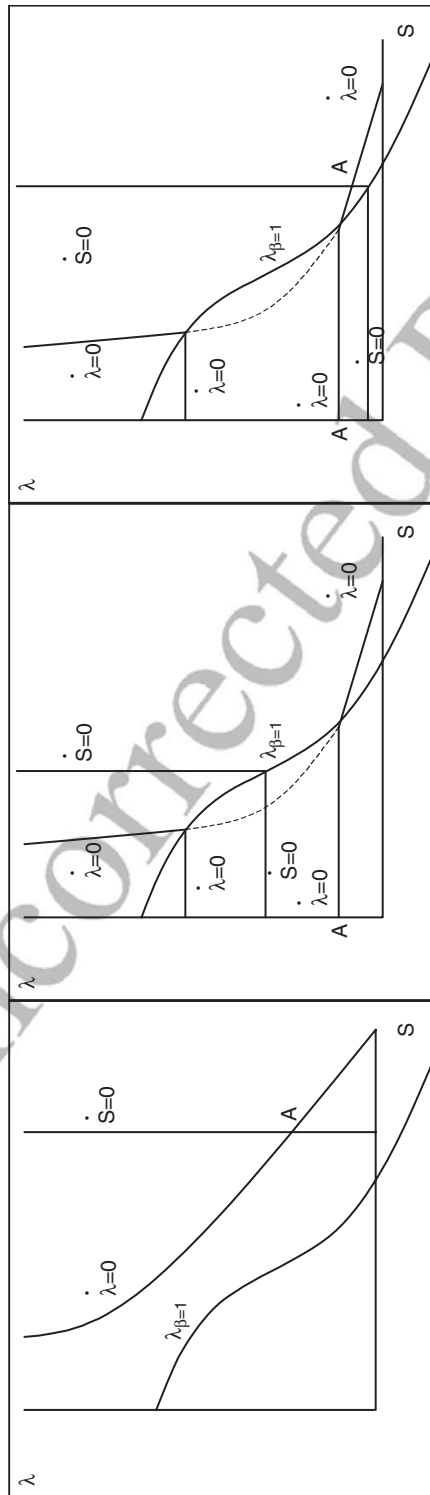
Proposition 1. *Either (1) There exists a unique steady state in which all production occurs using the modern technology, or (2) There exists a unique steady state in which all production occurs using the traditional technology, or (3) There exist two steady states, one of which has all production occurring with the modern technology and one of which has all production occurring with the traditional technology.*

Somewhat surprisingly a hybrid steady state where both modern and traditional production methods exist side by side is not a possibility. However multiple steady states are a possibility. The set of possible steady state configurations are depicted in the panels of Figure 3 below, with steady states denoted by capital A.

Figure 4 depicts the phase diagram and transitional dynamics of the steady state corresponding to the situation where neither locus crosses $\lambda_{\beta=1}$. Here the unique steady state of the system is the point A where traditional production stops entirely. An economy starting without skills in the modern technology will eventually develop a positive steady state skill level, and full modern production. In the example depicted in Figure 4, the unique steady state is characterized by a positive shadow value for managerial skills (since the steady state value of λ is positive). This is due to the fact that the vertical $\dot{S} = 0$ line is to the left of S' . If instead the $\dot{S} = 0$ line were to the right of S' , there would be no managerial skill premium in the steady state. This case is depicted by the dotted line for $\dot{S} = 0$ in the figure, and it intersects with the $\dot{\lambda} = 0$ line at a point where λ is equal to zero. In this case, managerial skills are accumulated to the point of abundance, S' , with the returns to the skill being fully dissipated in the resulting steady state. Note that such dissipation of the returns to knowledge arises even though there are no market imperfections, which contrasts with most models of knowledge creation and dissipation which rely on market imperfections.

Figure 5 depicts the phase space when both loci cross the $\lambda_{\beta=1}$ line. Here again, there is a unique steady state reached independently of the economy's starting position. Now, however, the economy's steady state sees it remain fully mired in the traditional technology. Though there is a positive managerial skill premium in the steady state ($\lambda > 0$), no one has incentive to accumulate skills in the modern technology. An economy inheriting a positive skill level will see it eventually approach zero along the transition path, since the endogenous creation of new managers does not offset the exogenous turnover rate. In such an economy, skill premia in the modern technology steadily rise as it approaches steady state. The agents in this economy choose not to sustain modern production because δ is too high relative to the desired investment in learning. Therefore, any stock of skill that is exogenously created, say through an explicit government training program or through migration from abroad, will eventually be depleted, causing a collapse in the use of the modern technology. Starting at a point on the transition path which is close to S' , the economy will have all production utilizing the modern technology. Even though the shadow value of skills is rising on this path, and all individuals are being

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48



C Modern and Traditional Steady State

B Unique Traditional Steady State

A Unique Modern Steady State

FIGURE 3

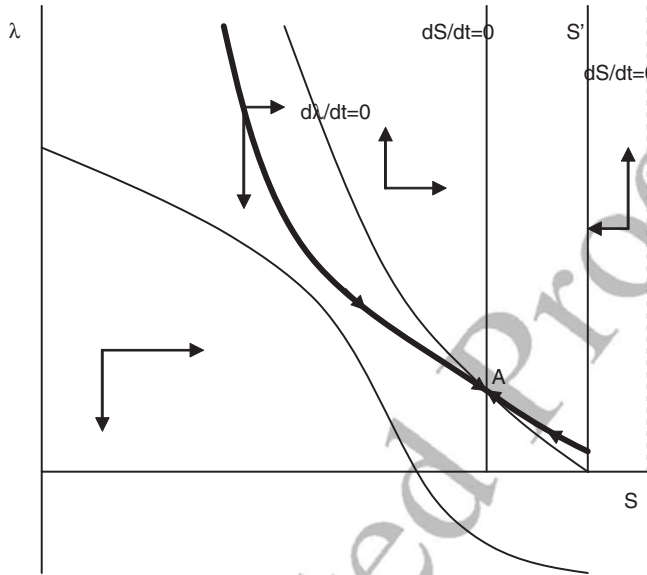


FIGURE 4

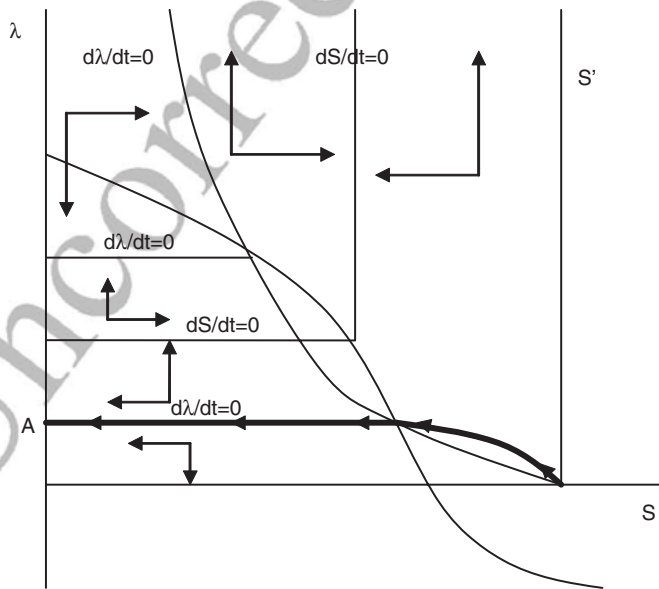


FIGURE 5

44 placed into learning the new technology, the economy's skill level is falling. Eventually, the
 45 transition path enters the $\beta < 1$ region, since the lower numbers of skilled managers would
 46 lower the wages of the unskilled below opportunity cost if all were placed into learning the new
 47 technology. When this occurs, the erosion of the skill stock is even greater and the economy
 48 converges along the $\dot{\lambda} = 0$ locus towards the steady state at A.

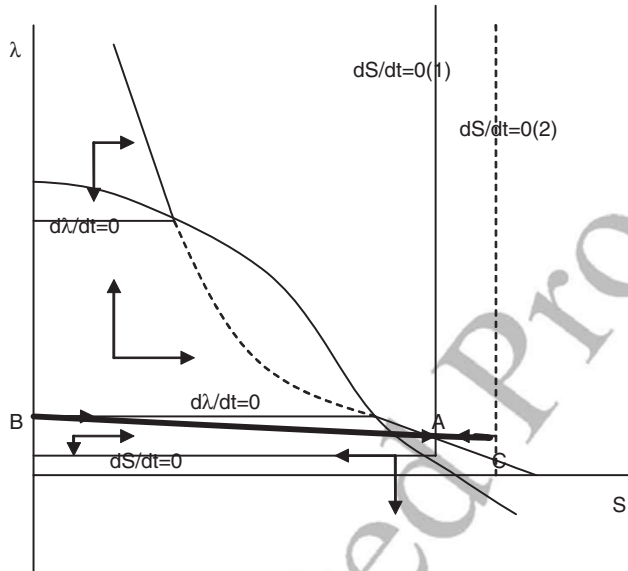


FIGURE 6

Figure 6 depicts the situation that occurs when only the $\dot{\lambda} = 0$ locus cuts the $\lambda_{\beta=1}$ line, or where if the $\dot{S} = 0$ locus also intersects, it does so at a value of S to the right of the last intersection point between $\lambda_{\beta=1}$ and $\dot{\lambda} = 0$. Here, in addition to the stable steady state, point A, where skills are accumulated and the economy is fully modern, there exists an unstable steady state at $S = 0$, the point labeled B. An economy starting at B would remain in the traditional technology because the shadow value of skills, λ , is not sufficiently high to warrant the sacrifice of earnings in the old technology in order to learn the new. However, the introduction of even an arbitrarily small number of skilled managers would lead to production with the modern technology. This is because management skills are accumulated as a by-product of production. With some skilled managers present, returns to working as unskilled in the new technology are extremely high, so that some unskilled workers arbitrage this by working there. But now, since δ is relatively low in this case, the number of managers increases, and so too does modern production. This induces more unskilled workers to work with the modern technology and learn managerial skills, further increasing their numbers until eventually all production is modern, and the point A is reached. Note that in this figure there is a qualitatively different case depicted for the dashed $\dot{S} = 0$ line denoted $\dot{S} = 0(2)$. This occurs when the $\dot{S} = 0$ locus does not intersect the $\lambda_{\beta=1}$ line.¹⁶

It is worth briefly discussing the nature of the steady state at point B in Figure 6. Although this steady state is unstable, it should not necessarily be considered irrelevant. For one, this steady state is not of the same type as many other examples in economics where a steady state at zero exists simply due to the assumption that accumulation cannot start without a positive initial amount of the accumulating factor. Here, even with zero skill, since $\Omega(0, U)$ is assumed greater than zero, skill can be accumulated if time is devoted to the modern technology. For

16. Here the horizontal arm of the $\dot{S} = 0$ line does not exist, but a similar steady state to A, denoted C, would ensue. In this later case, the steady will be characterized by the absence of skill premia if the $\dot{S} = 0$ locus is to the right of S' .

1 this reason a steady state with zero skill does not always exist, as for example in Figure 4.
 2 The steady state at a point such as B in Figure 6 arises only when $\dot{S} = 0$ at $S = 0$ becomes an
 3 equilibrium response. At point B, although an ϵ perturbation of S could trigger the accumulation
 4 of skill, the economy would stay an arbitrary long time near the no skill steady state if the ϵ
 5 perturbation is sufficiently small. For this reason, a steady state of the type illustrated by point
 6 B in Figure 6 may have some relevance for understanding the development process, which for
 7 some countries has been characterized by extremely slow take-up of improved technologies.

8 As the analysis above shows, the steady state configuration governing the economy's
 9 dynamics depends critically on the relative locations of the loci sketched above. Since these are
 10 determined by the values of exogenous parameters, δ and ρ , we can characterize the mapping
 11 between these values and eventual outcomes.

12
 13 **Proposition 2.** *For a given productivity level in the modern technology, F , and a given*
 14 *managerial learning technology Ω the $\rho - \delta$ space can be divided into three regions*
 15 *(a) For low values of ρ and δ , denoted region A, the economy converges to a unique steady*
 16 *state with all production occurring in the modern technology.¹⁷ b) For high values of ρ and δ ,*
 17 *denoted region B, the economy converges to a unique steady state with all production occurring*
 18 *in the traditional technology.¹⁸ c) For intermediate values of ρ and δ , denoted region C, if the*
 19 *economy starts with $S = 0$, it remains there with all production occurring in the traditional*
 20 *technology. If however the economy starts with $S > 0$, it converges to a steady state with all*
 21 *production occurring in the modern technology. Furthermore, there exists a δ^* , such that if*
 22 *$\delta < \delta^*$, then there is no managerial skill premium in a steady state where all production occurs*
 23 *with the modern technology.¹⁹*
 24

25 Figure 7 summarizes this proposition's implications for possible steady state outcomes.
 26 Regions A, B, and C correspond to the configurations in Figures 4, 5, and 6 respectively.²⁰

27 In the lower left corner of Figure 7 (i.e. low values of δ and ρ) lies the region where
 28 an economy with an arbitrary level of initial skills (including zero) will fully transform to
 29 modern production. If in this case $\delta < \delta^*$, there will be no managerial premium in the steady
 30 state. The disappearance of a managerial premium follows directly from the low labor turnover
 31 that ensures steady state skills are high. For higher values of δ , while staying in Region A,
 32 the modern technology will always fully develop, but the higher values of δ will allow the
 33 persistence of a skill premium in the steady state. For high values of both δ and ρ , as in
 34 Region B, development of the modern technology is not sustainable. An economy inheriting
 35 some skilled managers, or receiving in-migration of them, would utilize the technology, and
 36 would also train some further managers in its use. However, the low rate of skill creation is
 37 not sufficient to offset the high labor turnover and therefore such an economy will eventually
 38

39 17. This region is defined by a critical value, denoted k^* , such that if $\rho + \delta < k^*$, all production eventually
 40 occurs using the modern technology

41 18. This region is defined by a critical value of δ , denoted δ^{**} such that for $\delta \geq \delta^{**}$, there exists a corresponding
 42 value of ρ , denoted ρ^* such that, if $\rho > \rho^*$ the economy converges to a unique steady state with all production
 43 occurring in the traditional technology.

44 19. In this case, the steady state value of λ is zero.

45 20. The point δ' where the lines dividing the two regions intersect is effectively the value of δ at which the
 46 modern technology will not produce more than the traditional technology in steady state. This can occur even though,
 47 by construction, the modern technology is more productive than the traditional at full abundance. The reason is
 48 because, at values of δ beyond δ' , the steady state of an economy implementing the modern technology would be so
 much lower than full abundance of managerial skill that the economy produces less with it than with the traditional
 technology.

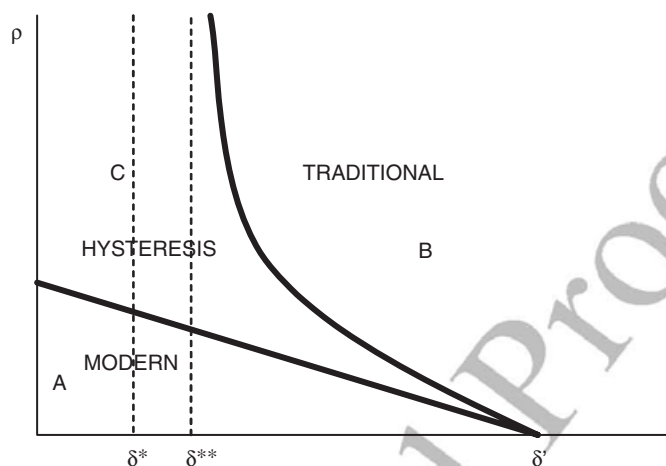


FIGURE 7

see such skills vanish from the population as the economy reverts back to full use of traditional technology. The middle region, region C, exhibits a knife edge type of hysteresis. The economy's low valuation of the future (high ρ), implies that without skills originally present, no unskilled worker would find it worthwhile to incur currently low productivity in order to accumulate managerial skill. However, with even a small amount present, they will be used in production, and since skills in the population depreciate slowly, since δ is relatively low, this will eventually lead to their diffusion through the population. The final outcome is a full transition to the modern technology.

2.6.1. Discussion. The hysteresis case in Figure 6 where a small "seeding" of skilled individuals leads to eventual diffusion of these skills throughout the sector, may appear to provide an explanation of the Bangladeshi (garment) and Indonesian (lumber) examples in the introduction. In both cases, skilled workers were transplanted into the poor country—by training and repatriating locals in the Bangladeshi example, and by transplanting skilled Koreans in the Indonesian example. The diffusion of their skills, through on-the-job learning, to future entrepreneurs rapidly lead to large follow-on effects. However, we do not think these examples are cases of the hysteresis regime. Instead, we believe the case which corresponds best to the anecdotes with which this paper starts is the one depicted in Figure 4. The reason is that this is the only case where returns at inception of the technology are high, and hence where there will tend to be rapid diffusion of skills and a dramatic mushrooming of the modern technology in the sector. In this case, as the transition path in Figure 4 suggests, all unskilled are willing to work in the modern sector in order to acquire skills, as the shadow value of managerial skills is extremely high, so that take-up is rapid. In the hysteresis case of Figure 6, by contrast, returns to skill at the start are low, and the initial movement of the unskilled to work with the small number of skilled in the modern technology is small. We return further to the model's implications for diffusion through migration in the next section.

3. MIGRATION INCENTIVES

Theories of cross-country growth differences that have attributed significant roles to embodied skills generally imply migration incentives that are perverse in comparison with actual migration

1 flows. This arises for standard neo-classical reasons.²¹ Since skills are relatively abundant in
 2 developed countries, returns should be low there, and skilled individuals should want to migrate
 3 to less developed countries where they are scarce. In the present framework, skills are embodied
 4 and fully rewarded, suggesting that a similar set of incentives should be in place. However,
 5 this is not the case.

6 Consider migration incentives between two countries with access to the same set of
 7 technologies, modern and traditional, but differing in one of three ways that will generate
 8 outcome differences. The first two sources of difference are fundamentals: the discount factor,
 9 ρ ; and the population turnover rate, δ . As we have seen previously, countries with low values
 10 of each of these are more likely to converge to steady states where the modern technology
 11 dominates. The final case is a difference implied by hysteresis, illustrated in Figure 6, where
 12 fundamentals are identical.

13 We focus on instantaneous incentives for migration. If a wage payment is higher—for a
 14 given skill level—in country i versus country j , we will say that a skilled worker has an
 15 incentive to migrate from country j to country i . We frame our discussion primarily around
 16 the incentive for skilled managers to migrate from richer to poorer countries, but the flip-side
 17 of this is an explanation of the brain-drain of skilled workers from poor to rich countries. We
 18 omit discussion of the migration incentives for unskilled workers since they are quite standard
 19 in the model in the sense that there is never an incentive for an unskilled worker to migrate
 20 from a rich country to a poor country.

21

22 3.1. Migration incentives due to differences in ρ

23

24 Consider two countries, rich, r , and poor, p , with identical values of δ but with differing
 25 discount rates such that $\rho^r < \rho^p$. The values of the respective ρ are such that the rich country
 26 is in Region A in Figure 7, while the poor country is in Region B.

27

28 The poor country is in a steady state where all production occurs in the traditional
 29 technology, and the rich is in one where all production is modern. In the rich country's steady
 30 state there is ongoing training of individuals in managerial skills, whereas in the poor country's
 31 steady state, managerial skills are not being produced. This does not mean that managerial skills
 32 would not be valuable in the poor country. Specifically, if one skilled manager from the rich
 33 country, where skills are widespread, were to migrate to the poor country, he would be able
 34 to set up production and utilize the modern technology. Moreover, the unskilled workers he
 35 employed would also acquire valuable managerial skills and therefore be willing to “pay” for
 36 these by working at a wage below the wage prevailing with the old technology. This would
 37 generate a positive return to managerial skill and would appear to imply strong incentives for
 38 migration of skilled individuals from the rich country to the poor one. This is not the case.²²

39

40 **Proposition 3.** *Income differences generated by differences in ρ across countries do not
 41 generate incentives for managers in a richer country to migrate to a poorer country.*

42

43 How can it be that wages are higher for managers in the rich country when managers are
 44 more abundant there? The answer is the teaching value of managerial skills. Recall that wage

45

46 21. This view is summarized in Romer's (1995) critique of Mankiw's (1995) emphasis on formal training as an
 47 explanation for cross-country differences.

48 22. The managerial wage in the poor country is given by $w^s(\delta) = F_1(Z, 1) + \lambda \tilde{\Omega}_1(Z, 1)$, where λ is given by

$$\lambda = \frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)}$$
 and where Z is the smallest root of $\frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)} = \frac{\bar{w} - F_2(Z, 1)}{\Omega_2(Z, 1)}$.

1 payments to a manager are made up of two components: the marginal product in production
2 and the teaching value. The lower value of ρ , which makes the country richer, does this
3 by inducing more individuals to accumulate managerial skills. This ensures a high teaching
4 value of managers and since many more unskilled workers are willing to learn managerial
5 skills, this also guarantees overall managerial productivity is higher there too. So, the rich
6 country's relatively high valuation of the future induces relatively many of its unskilled to
7 learn managerial skills, making both the teaching value of managerial skills and the marginal
8 product of managers high. The empirical implication of this is that, in the rich country, each
9 manager should be observed to work with relatively more unskilled workers than in the poor.

10 This finding fits precisely with the empirical evidence presented on manager/worker ratios
11 across countries in Acemoglu and Newman (2002). They report averages of the ratio of
12 managerial to production workers in six OECD countries, all of which are below 25%. In
13 contrast, the corresponding ratio in Sub-Saharan Africa averages 41%. Our explanation for this
14 is that the lower steady state value of managerial skills (due to the higher rate of discount, ρ)
15 in the Sub-Saharan economies implies workers are less willing to pay the cost of being a low
16 productivity worker required to obtain them.

17 The reciprocal of this is the existence of strong incentives for a brain-drain induced by
18 cross-country differences in ρ . Skilled individuals from poor countries, where skills are scarce,
19 will strictly prefer to migrate to rich countries where their skills are relatively abundant because
20 by doing so they will work with more unskilled subordinates (reflecting the higher teaching
21 value of their skills) and have a higher marginal productivity. As the empirical record of
22 migrations shows, such incentives appear to be strongly in place, and to be consistent with the
23 managerial/worker ratios reported above.

24 A corollary to the result above follows when considering changes that would occur when
25 there is an improvement in the modern technology so that a country that previously did not find
26 it worthwhile to invest in the modern technology, now does so. Technical change, for instance of
27 the form which would imply an improved production function, \widehat{F} , which everywhere dominates
28 production function F would lead to the following:

29
30 **Corollary 1.** *An increase in the productivity of the modern technology strictly increases*
31 *the skill premium and (weakly) increases skill supply.*
32

33
34 The corollary is thus consistent with an observation that has been documented in the
35 skill-biased technical change literature. Increases in the productivity of a technology, or the
36 introduction of a new technology, is correlated with increased skilled supply in that technology,
37 and increases in the skill premium to using it. Many microeconomic studies have documented
38 this; see for example Autor, Levy and Murnane (2003), and a survey of the skill biased technical
39 change literature in Acemoglu (2002).
40

41 3.2. Migration incentives due to differences in δ

42
43 With an international capital market that is able to equalize market discount rates across
44 countries, individuals will effectively behave as if they have de facto identical discount rates.²³
45 Though one may doubt the realism of such smoothly functioning international capital markets,
46 it is worthwhile understanding whether the strong migration result above depends critically
47

48 23. We are grateful to the editor for pointing out this possibility.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

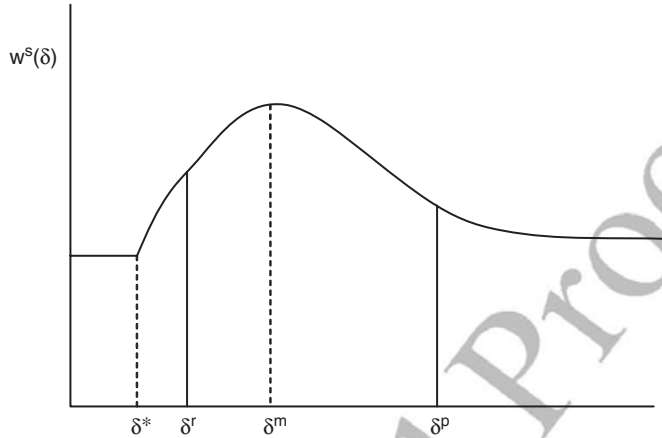


FIGURE 8

on capital market imperfections. The answer is no, it is possible to obtain similar results for differences in δ , though now some caveats apply.

Consider the induced migration incentives between two countries that differ in terms of δ , but have the same value of ρ . The first country has a high value of δ , denoted δ^p , placing it in region B of Figure 7 which leads it to use only the traditional technology in steady state. The second country has a lower value of δ , denoted δ^r , placing it in either region A or C of Figure 7.²⁴ In its steady state, all individuals work with the modern technology. Incentives for skilled migration generated by these differences in δ are more complicated than in the ρ case because now both the $\dot{\lambda} = 0$ and the $\dot{S} = 0$ schedules are altered. In order to understand these, in Figure 8 we plot the steady state wage of skilled workers as a function of δ . Recall that their wages can be expressed as follows $w^s(\delta) = F_1(Z, 1) + \lambda \tilde{\Omega}_1(Z, 1)$, where Z is the ratio of managers to unskilled workers in the modern technology. As δ changes, wages of managers change due to changes in both Z and λ . For the case where δ is in region B, there are no skilled managers in the steady state, which may give the impression that Z , and hence w^s , are not well defined. However, this is not the case since, as seen in Figure 5, there is a well defined steady state value for λ ²⁵ and this value of λ pins down the relevant value of Z by the optimality of the β choice when $\beta < 1$. This condition can be expressed as $\lambda = \frac{\bar{w} - F_2(Z, 1)}{\Omega_2(Z, 1)}$. For the case where the steady state is characterized by exclusive use of the modern technology, then the relevant values for Z and λ are determined by the $\dot{S} = 0$ and $\dot{\lambda} = 0$ conditions when $\beta = 1$.

The steady state relationship between w^s and δ has an inverse U shape. The turning point value δ^m is the δ given by the boundary between regions C and B in Figure 7. Accordingly, to the right of δ^m , the relationship between w^s and δ is negative and a country in this region is underdeveloped. The skilled wage is decreasing in δ in this region because higher turnover implies the unskilled are less willing to 'pay' to learn managerial skills through low wages. To the left of δ^m , the country is characterized by exclusive use of the modern technology and is always richer (in terms of output-per-capita) than countries with $\delta > \delta^m$. Here, neo-classical

24. If a country has a parameter configuration which places it in Region C of Figure 7, we consider here that such a country is in the high output steady state.

25. This λ is given by $\lambda = \frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)}$, where Z is the smallest root of $\frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)} = \frac{\bar{w} - F_2(Z, 1)}{\Omega_2(Z, 1)}$.

1 forces dominate so that wages for the skilled in steady state are lower the greater the proportion
2 skilled, i.e., for lower δ . The non-monotonic relationship between w^s and δ implies that a
3 manager in a developed/rich country may not want to migrate to a poor country where his
4 skills are less abundant. In fact, for any given poor country, there exists a range of δ s (with
5 $\delta < \delta^m$) where rich countries will exclusively use the modern technology in steady state, and
6 where the skilled will not want to migrate to the poor country. This observation is summarized in
7 Proposition 4. Once again, the reason why a manager in a rich country may not want to migrate
8 to the poor country is because the teaching value of having skill is greater in the rich country.

9
10 **Proposition 4.** *Cross-country income differences generated by differences in δ generate*
11 *incentives for skilled managers to migrate to poor countries only if δ in the rich country is*
12 *sufficiently low.*

13
14 As distinct from the ρ case, there are now cases when δ in the rich country is sufficiently
15 low where skilled managers will wish to migrate to the poor and set up the modern technology
16 there. But these skilled migrations do not have the ability to explain the examples characterized
17 by rapid diffusion of modern methods with which this paper started. These rapid diffusion
18 anecdotes lead to the diffusion and dominance of the modern technology through expansion
19 of domestic managerial capacities. Migration incentives due to differences in δ do not imply
20 this if the receiving country is in region B. What they imply instead is the emergence of a
21 dependence relationship. Even if some local managerial skill is developed in the poor country,
22 it is not sufficient to sustain the expansion of the modern technology, and reliance on foreign
23 skill imports is maintained. A country with parameters in region B would need to continually
24 import managerial skill from abroad to support the new technology since it will not generate
25 enough domestic managerial skills on its own.

26 It may seem odd to treat as country level primitives the parameters ρ and δ when allowing
27 individuals, who embody such primitives, to relocate. Presumably, if enough individuals from
28 the low ρ or δ country migrate to the high one then there is a force for homogenization of
29 these values. However this changes nothing in our propositions 3 and 4 above. These explain
30 why skilled managers will *not* find it worthwhile to migrate from countries where skills are
31 abundant to ones where they are scarce. Since they explain cases under which there will be no
32 migration, there is no force operating to homogenize these underlying parameters.

33 3.3. Differences due to hysteresis

34
35 Now consider two economies with identical ρ, δ values that lie in the hysteresis area between
36 the bold lines in Figure 7. Country r is at a steady state like that sketched at point A in Figure 6,
37 while the poor country p is at a steady state corresponding to no skill acquisition—point B in
38 Figure 6. In this case, the skilled have a strict incentive to migrate to the place where skills
39 are scarce.

40
41
42 **Proposition 5.** *When cross-country differences are driven by hysteresis and two countries*
43 *with identical fundamentals have different levels of development, there is incentive for skilled*
44 *individuals to migrate from the rich to the poor country.*

45
46 Here, the migration incentives have a purely neo-classical flavour. The country abundant
47 in skilled managers will see out-migration to the otherwise identical country which has not yet
48 developed. Moreover, the introduction of a small number of skilled managers into the poor

1 country will see it transition along the path in Figure 6 from the point B to eventually reach
 2 the same level of output per capita as the rich country, with full use of the modern technology.

3 Some aspects of migration incentives generated in this case are consistent with the
 4 transitions documented in the introduction. It can explain why migration of a small group
 5 of managers could start a process which would eventually lead to the dominance of the new
 6 technology through the development of local skill. However, it does not explain how the
 7 takeoffs could have been so rapid. As previously discussed, in the hysteresis configuration, a
 8 small injection of managers leads to a rather slow diffusion process as the incentives for skill
 9 accumulation are not very strong.

10 For this reason, we think the best way to understand the cases discussed in the introduction
 11 is that the environments in the receiving countries in each case were likely changing,
 12 corresponding to movement out of region B . These could have been due to policy incentives,
 13 health improvements, or any other factors affecting future valuations and turnover. These
 14 changes explain both the emergence of migration incentives and rapid diffusion of new
 15 technologies by expansion of a domestic managerial class. This interpretation suggests that
 16 migration incentives emerged as a result of the changes internal to the less developed country,
 17 that migration was not the key driving force but simply sped up the process.

18 19 4. IMPLICATIONS FOR GROWTH ACCOUNTING 20

21 We now consider whether the process of managerial skill acquisition we have presented is
 22 of quantitative relevance in the development process. Can it potentially explain a substantial
 23 fraction of cross-country differences in income? At a first pass, the answer from the growth
 24 accounting literature appears to be a clear no. This literature (see for example Klenow and
 25 Rodriguez-Clare (1997)) finds that skills, whether acquired on-the-job or by schooling, account
 26 for only a small fraction of income differences across countries. In order to evaluate the role
 27 of on-the-job skill acquisition, this literature uses estimates of the returns to job experience to
 28 calculate the share of income paid to on-the-job acquired skills. This share of income is found
 29 to be small from which it is generally inferred that no element of on-the-job skill acquisition
 30 can explain an important fraction of cross-country income differentials. We argue here that
 31 such an inference would not follow if the data were generated by the type of managerial
 32 skill acquisition process we have presented. We start by examining what our model implies
 33 regarding standard growth accounting exercises, under the assumption that returns on-the-job
 34 are estimated using experience profiles. We illustrate why standard growth accounting exercises
 35 would not properly assess the role of on-the-job learning if the data were generated by our
 36 model. Since the impact upon growth accounting depends upon whether a cross-section of
 37 countries or a time-series is considered, we explore both cases.

38 39 4.1. *Returns to experience in a cross-section* 40

41 Growth accounting in a cross-section involves decomposing differences in (log) income-per-
 42 worker into components associated with factor accumulation and a residual. For example,
 43 consider the following three factor decomposition: $\ln \frac{Y_i}{L_i} = \alpha_i \ln \frac{EXP_i}{L_i} + \gamma_i \ln \frac{X_i}{L_i} + \ln A_i$, where
 44 $\frac{EXP_i}{L_i}$ is experience per worker in country i , α_i is the share paid to experience, X_i is some
 45 composite of physical and human capital, (usually entering separately, but bundled together
 46 for simplicity here) together with its share, γ_i and A_i is the residual, often referred to as TFP.
 47 The term α_i can be computed using estimates from a Mincer type earnings regression based
 48 on micro data. In explaining cross-country differences in output per capita say between two

1 countries i and j , the term $\frac{\ln A_i - \ln A_j}{\ln \frac{Y_i}{L_i} - \ln \frac{Y_j}{L_j}}$, is usually interpreted as the contribution of factors that
 2 have not been purposefully accumulated. This is because it is assumed that the contribution of
 3 differences in skills learned on-the-job is reflected in the returns to the experience. In order to
 4 see how this may be misleading, now suppose that countries i and j have different values of
 5 δ and ρ , and that the growth process is driven by on-the-job managerial skills acquisition as
 6 we have modeled it. Assume for now that $\delta_i \leq \delta^*$. Furthermore, assume that country i has a
 7 low value of ρ so that it lies in Region A in Figure 7, whereas country j has a high value of
 8 δ and ρ so that it lies in region B.

9
 10 Consider an accounting exercise exploring level differences in income between countries
 11 i and j using observations drawn from the steady state. If generated by our model, income
 12 per-capita is higher in country i because it has more managerial skills, which have been
 13 accumulated through the use of the new technology. However, returns to work experience
 14 estimated from micro data are precisely zero in country i . This is because managerial skills are
 15 abundant in a steady state with $\delta_i \leq \delta^*$ and hence have a zero return. Consequently, all of the
 16 difference in income levels caused by differences in on-the-job learning will be picked up by
 17 the residual. This would lead researchers to incorrectly interpret the cause of the difference in
 18 income per capita between i and j as being due to a disembodied factor, such as differences
 19 in technological opportunities.

20 Note that the same over-attribution of differences to the residual would also occur
 21 were country i to be located in the region A where $\delta_i > \delta^*$. Here, experience differences
 22 would now be measured to pick up some of the difference in output per worker, but this
 23 measured contribution need bear no relationship to the true contribution. In fact, the degree of
 24 understatement would be inversely related to the size of δ .

25 4.2. Returns to experience in a time series

26
 27 Growth accounting can also be performed using time series data. Consider a single country
 28 that is experiencing growth in output per capita according to our process of managerial skills
 29 acquisition. In attempting to pick up the contribution of various factors to the country's growth
 30 over time, one would generally perform the following decomposition: $\frac{\dot{y}}{y} = \alpha_t \frac{\text{exp}}{\text{exp}} + \gamma_t \frac{\dot{x}}{x} + \frac{\dot{A}}{A}$,
 31 where x , y and exp are all in per-capita terms, and A represents the TFP residual. Skills learned
 32 through experience are usually proxied as a weighted average of the economy's age structure,
 33 or a weighted average of time in the work-force. In our example, economy i experiences growth
 34 in output per capita all the way through its transition due to an increase in the fraction of the
 35 population which has managerial skill. But the age of workers, and hence measured experience,
 36 is unchanged through the transition. Thus proxying experience by the age structure or time in
 37 the labor force would, in data generated by our model, lead to the finding that $\frac{\text{exp}}{\text{exp}} = 0$. The
 38 outcome is therefore the same as in the cross-section: a growth accounting exercise performed
 39 on time series data generated by our model will attribute all of the on-the-job learning to the
 40 residual term, $\frac{\dot{A}}{A}$. Income growth is again erroneously attributed to an unexplained residual
 41 factor instead of its true source; managerial skills purposefully accumulated on the job.

42 4.3. Correcting growth accounting exercises

43
 44
 45 In the presence of managerial skill acquisition (of this learning-by-seeing type), standard growth
 46 accounting will give misleading estimates of the contribution of disembodied technology to
 47 cross-country or cross-time differences in output-per-worker. Here we illustrate how growth
 48 accounting can be modified to deliver a meaningful decomposition. The first requirement for

1 this is that the fraction of managers in the population be observable, and hence this is what
2 we will assume.

AQ9 Let us focus on the simplest case where output in country i can be expressed as
3 a function of managerial skill S_{it} according to a CRS production function of the form
4 $\frac{Y}{L_{it}} = y_{it} = A_{it} F\left(\frac{S_{it}}{L_{it}}, 1 - \frac{S_{it}}{L_{it}}\right) = A_{it} F(s_{it}, 1 - s_{it})$, where A_{it} is the index of productivity
5 for country i at time t , s_{it} is the fraction of the population that are managers and $F(\cdot)$
6 is assumed to be unknown. Taking the time derivative of this relationship implies $\frac{\dot{y}_i}{y_i} =$
7 $\theta(t)_i \frac{\dot{s}_i}{s_i} + \frac{\dot{A}_i}{A_i}$, where $\theta(t)_i$ is the share of income paid to managerial skill and is given by
8 $\theta(t)_i = \left(\frac{s_{it} F_1(s_{it}, 1 - s_{it})}{F(s_{it}, 1 - s_{it})}\right) \left(1 - \frac{F_2(s_{it}, 1 - s_{it})}{F_1(s_{it}, 1 - s_{it})}\right)$. Note that $\theta(t)_i$ will be equal to zero if managerial
9 skill is abundant. One can then obtain the (log) index of productivity for country i at time
10 t as follows: $\ln A_{it} = \ln y_{it} - \int_{-\infty}^t \theta(\tau)_i \frac{\dot{s}}{s} d\tau = \ln y_{it} - \theta(t)_i \ln s_{it} + \int_{-\infty}^t \frac{\partial \theta(\tau)_i}{\partial \tau} \ln s_{it} d\tau$. This
11 expression makes it immediately clear that the contribution at time t of disembodied technology
12 in output-per-capita differences cannot be measured using only time t information, unless $\theta(t)_i$
13 does not change over time. However, if managerial skill is being acquired over time, then $\theta(t)_i$
14 will necessarily be declining over time. It is therefore necessary to use the full time path of
15 returns to managerial skill to properly assess the role of accumulated factors versus disembodied
16 technology in differences in output-per-capita. Using only $\theta(t)_i \ln s_{it}$ as the contribution of
17 managerial skills will lead to an over-estimation of the role of disembodied technology in
18 cross-country income differences.
19
20

21 The argument can be easily extended to allow for multiple vintages of technology. The
22 main insight remains that in the presence of managerial skill acquisition, the share of income
23 attributed to managerial skill within each vintage will decline with adoption, even if total
24 returns to managerial skill do not decline because of the arrival of successive vintages. In such
25 cases, it is necessary to use the whole history of returns and quantities to evaluate the role of
26 such on-the-job accumulated skills.
27
28

29 4.4. *A quantitative example of the role of managerial skill in cross-country income differences*

30 In this section we explore how big the underestimate of managerial skill acquisition's role could
31 be. Suppose we have two countries with access to the same technologies but located in different
32 regions of Figure 7. We investigate whether this can lead countries to have quantitatively large
33 differences in income-per-capita and whether such countries can exhibit large differences in
34 TFP, as traditionally measured. This amounts to examining whether small differences in ρ or
35 δ can cause countries to exhibit large differences in output-per-capita and measured TFP. To
36 do this, we first calibrate our two technology model.

37 The approach we adopt for calibration is as follows. First, we normalize the level of
38 output-per-worker using the old technology to be equal to 1. Then we assume that the modern
39 technology takes the form $\Gamma S^\alpha U^{1-\alpha}$, and we assume that the learning process is driven by a
40 fixed learning rate Ω that is independent of S and U . As will become clear, we will choose α
41 and Ω such that a country which chooses to adopt the modern technology will exhibit a wage-
42 experience profile that is similar to that observed in the US. We also assume that countries
43 have the same rate of time preference, which will either be 5% per year or 8% year per. This
44 leaves us with three parameters: Γ , δ_1 and δ_2 , where δ_1 and δ_2 represent the death rates in our
45 two countries denoted 1 and 2. One possibility would be to choose an arbitrary value for Γ and
46 explore how small differences in δ translate to differences in output-per-capita. However, the
47 answer to such an exercise is sensitive to the choice of Γ , since results will change substantially
48 depending on whether or not Γ is such that small differences in δ cause the countries to be in

different regions of Figure 7. Since the choice of Γ in this case is arbitrary, we instead proceed as follows. We set an identical value for δ in the two countries, and this is set so that average life expectancy is 55 years ($\delta = \frac{1}{55}$). Then we choose Γ such that the economies are either on the line that separates region A and C, or on the line that separates regions B and C. These Γ s are given by the following functions of model parameters.²⁶

Given values for $\rho, \delta, \Omega, \alpha$, the Γ that places an economy on the line that separates regions B and C in Figure 7, denoted Γ_{B-C} , is:

$$\Gamma_{B-C} = \frac{\delta + \rho + \Omega}{\left(\frac{\Omega}{\delta^A}\right)^\alpha (\delta + \rho (1 - \alpha))}. \quad (13)$$

Similarly, given values for $\rho, \delta, \Omega, \alpha$, the Γ that places an economy on the line that separates regions A and B in Figure 7, denoted Γ_{A-B} , is:

$$\Gamma_{A-B} = \frac{\delta + \rho + \Omega}{(\delta + \rho) (1 - \alpha) \left(\frac{\Omega}{\delta + \rho}\right)^\alpha + \Omega \alpha \left(\frac{\Omega}{\delta + \rho}\right)^{\alpha-1}}. \quad (14)$$

Only at precisely these two Γ values is the economy at a knife-edge where both full adoption of the modern technology is a steady state, and zero adoption of the modern technology is a steady state. In the first case, a slightly higher value of δ causes no adoption to be the only steady state (as in Figure 5), while a slightly lower value of δ leads to hysteresis—either full or no adoption to be a steady state (as in Figure 6). In the second case, a slightly higher value of δ also causes the economy to be in the hysteresis region (as in Figure 6), while a slightly lower value leads to the adoption of the modern technology as the only steady state (as in Figure 4). For such parameter configurations, we examine the extent of income differences between the full adoption and the no adoption steady states. In other words, this approach illustrates the extent to which cross-country differences in income could be generated by infinitely small differences in work-life expectancy. Obviously, the effects we calculate for these cases would be further amplified by incorporating more heterogeneity in δ . However, since differences in work-life expectancy for young adults do not vary greatly across countries, we do not pursue this additional amplification mechanism. Note that the comparison using Γ_{BC} may be considered more conservative since it does not rely on a comparison with the no adoption equilibrium when it is locally unstable in the hysteresis case.

To perform the calibration exercise, the only important remaining element is the selection of values for α and Ω . Since skills are acquired on the job, neither of these parameters is directly observed, but we can calibrate these indirectly by matching the model's implications for wage experience profiles under the modern technology with empirical counterparts in US data. Specifically, the two pieces of information we utilize are i) the maximum

26. To compute Γ_{B-C} , we solve for the value of Γ such that the $\lambda_{\beta=1}$ line intersects the $\dot{\lambda} = 0$ line at the point $S = \Omega / (\Omega + \delta)$. Substituting in the Cobb-Douglas form of the production function and rearranging this tangency condition yields $\frac{\Gamma}{\bar{w}} = \frac{\delta + \rho + \Omega}{\Omega \alpha \left(\frac{S}{1-S}\right)^{\alpha-1} + (\delta + \rho)(1-\alpha) \left(\frac{S}{1-S}\right)^\alpha}$. Substituting for $S = \Omega / (\Omega + \delta)$ yields the equation for Γ_{B-C} above.

Γ_{A-B} is given by the Γ generating a tangency between the $\lambda_{\beta=1}$ locus and the $\dot{\lambda} = 0$ line. To compute this, first note that the distance between these lines is given by: $\text{Dist} = \frac{\alpha \Gamma (S/(1-S))^{\alpha-1} - (1-\alpha) \Gamma (S/(1-S))^\alpha}{\delta + \rho + \Omega} - \frac{\bar{w} - (1-\alpha) \Gamma (S/(1-S))^\alpha}{\Omega}$. It is easy to verify that the difference, Dist , is minimized at $S = \frac{\Omega}{\Omega + \delta + \rho}$, so that a tangency between these schedules occurs if and only if at $\frac{S}{1-S} = \frac{\Omega}{\delta + \rho}$ we have $\frac{\alpha \Gamma (\Omega / (\delta + \rho))^{\alpha-1} - (1-\alpha) \Gamma (\Omega / (\delta + \rho))^\alpha}{\delta + \rho + \Omega} = \frac{\bar{w} - (1-\alpha) \Gamma (\Omega / (\delta + \rho))^\alpha}{\Omega}$, which rearranges to yield the productivity differential $\frac{\Gamma_{A-B}}{\bar{w}}$ above.

1 amplitude of the wage-experience profile, which should reflect the manager/unskilled wage
 2 differential, denoted $\frac{w^S}{w^U}$, and ii) the rate of return to a worker's first year of experience,
 3 denoted $r(0)$.

4 Given the Cobb-Douglas specification, the steady state skilled/unskilled wage differential
 5 under the modern technology is equal to:

$$\frac{w^S}{w^U} = \frac{\alpha}{1-\alpha} \frac{\delta}{\Omega} \quad (15)$$

6
 7
 8
 9 Under a constant on-the-job skills learning technology, Ω , a worker of experience τ
 10 has expected wage given by $w(\tau) = (1 - \exp[-\Omega\tau]) w^S + \exp[-\Omega\tau] w^U$. It follows that
 11 the rate of return to increased experience at experience level τ is given by $\frac{d \ln w(\tau)}{d\tau} =$
 12 $\frac{\Omega \exp[-\Omega\tau] \left(\frac{\alpha}{1-\alpha} \frac{\delta}{\Omega} - 1 \right)}{(1 - \exp[-\Omega\tau]) \left(\frac{\alpha}{1-\alpha} \frac{\delta}{\Omega} + \exp[-\Omega\tau] \right)}$. At $\tau = 0$ the observed $r(0) = \frac{d \ln w(0)}{d\tau}$, then reduces to:

$$r(0) = \frac{\alpha}{1-\alpha} \delta - \Omega \quad (16)$$

13
 14
 15
 16
 17
 18 The solutions to (15) and (16) yield $\Omega = \frac{r(0)}{\frac{w^S}{w^U} - 1}$ AND $\alpha = \frac{r(0)/\delta \left(\frac{w^S/w^U}{w^S/w^U - 1} \right)}{r(0)/\delta \left(\frac{w^S/w^U}{w^S/w^U - 1} \right) + 1}$, so we can
 19
 20
 21 calibrate the remaining two parameters Ω and α .

22 The estimate of $\frac{w^S}{w^U}$ we use is obtained from Heckman, Lochner and Todd (2008), who
 23 report a maximal earnings gap between young and old workers in the vicinity of a factor of
 24 3 for the U.S. (see Figure 1 of their paper). We thus set $\frac{w^S}{w^U} = 3$, as our baseline. We then
 25 choose a value for the first year's experience $r(0) = 0.15$ as a baseline, also based on the same
 26 source. Since both of these measures are sensitive to numerous econometric and conceptual
 27 issues discussed at length by Heckman, Lochner and Todd, we report the sensitivity of our
 28 estimated productivity differentials for ranges of 30% above and below our baseline, i.e., for
 29 values of 2 and 4 for the wage differential, and for values of 10% and 20% for the first year
 30 returns to experience.

31

32

33

4.5. Results

34 We first compute the value of Γ for our benchmark calibration. We compute this for the case
 35 where the economies lie on the boundary between regions B and C, and for the case where
 36 economies are at the boundary of regions A and C. For the A and C comparison, we assume that
 37 an economy in region C is in a steady state with non-adoption of the modern economy. Since
 38 this outcome is locally unstable, we will refer to this comparison as the fragile comparison
 39 and the B and C comparison as the robust one. However, as we have indicated previously, we
 40 believe that a comparison with the locally unstable equilibrium is nonetheless interesting since
 41 it can be interpreted as examining an economy which if perturbed may take a very long time
 42 to converge to the full adoption of the modern technology.

43 The estimates of Γ we obtain for our benchmark calibration ($\rho = .05$, $\delta = \frac{1}{53}$, $\frac{w^S}{w^U} = 3$ and
 44 $r(0) = 0.15$.) are $\Gamma = 1.76$ for the robust case (boundary B-C), and $\Gamma = 1.92$ for the fragile
 45 case. These range from a low of 1.3 to a high of 2.8, for the robust case, and a low of 1.4
 46 to a high of 3.5 for the fragile case. Higher ρ values increase this differential. These numbers
 47 are extremely robust to the chosen values of δ . For example, in the baseline case with $\frac{w^S}{w^U} = 3$
 48 and $r(0) = 0.15$, almost doubling δ to 1/30 changes the value of Γ/\bar{w} in the fragile case from

1 1.92 to 2.08, while lowering δ to 1/80 lowers it to 1.80. We thus do not report separate tables
 2 by differing δ values as the quantitative estimates are largely unaffected.

3 With these Γ values we can now consider the attribution to factors and technology that
 4 would be implied by a growth accounting exercise. Since both countries in either region have
 5 access to the same set of technologies, the contribution that differences in TFP makes to
 6 explaining cross-country output differences, if correctly imputed, should be zero. We proceed
 7 by assuming the most favorable case for a growth accounting whereby managerial skill is
 8 fully observed: one can observe both the fraction of the population with managerial skill;
 9 which is equal to $\frac{s}{L} = \frac{\Omega}{\Omega+\delta}$ in the economy which uses the modern technology, and one
 10 can observe the fraction of total income that is a return to managerial skill acquisition;
 11 which is given by $\theta = \alpha(1 - \frac{(1-\alpha)\Omega}{\alpha\delta})^{27}$ in the full adoption steady state. However let us
 12 assume, following common practice, that cross-country growth accounting is done using only
 13 information at a point in time, and hence the contribution of managerial skill to log output-
 14 per-capita is evaluated as $\alpha(1 - \frac{(1-\alpha)\Omega}{\alpha\delta}) \ln(\frac{\Omega}{\Omega+\delta})$. In this case, the estimated % difference in
 15 TFP between two countries 1 and 2 would be calculated as $\ln \frac{Y_1}{L_1} - \theta_1 \frac{S_1}{L_1} - (\ln \frac{Y_2}{L_2} - \theta_2 \frac{S_2}{L_2})$.
 16 In the case, where country 1 is in a steady state using the modern technology, while
 17 country 2 is in a steady using the old technology (with output-per-capita normalized to 1),
 18 the TFP difference can be calculated as $\ln A_1 - \ln A_2 = \ln \frac{Y}{L_1} - \alpha(1 - \frac{(1-\alpha)\Omega}{\alpha\delta}) \ln(\frac{\Omega}{\Omega+\delta}) =$
 19 $\ln \Gamma + (1 - \alpha) \ln(\frac{\delta}{\Omega+\delta}) + \alpha(\frac{(1-\alpha)\Omega}{\alpha\delta}) \ln(\frac{\Omega}{\Omega+\delta})$.

20 The tables below indicate the imputed differences in productivity between countries, i.e.,
 21 the ratio $\frac{A_1}{A_2}$ in the two steady states calculated using this decomposition for the fragile and
 22 robust cases. The tables report this ratio for $\frac{w^S}{w^U}$ ranging from 2 to 4, for $r(0)$ ranging from
 23 0.1 to 0.2, and for ρ values of 0.5 and 0.8.

25 Inferred Productivity Difference Estimates, robust (B-C) case

$\frac{w^S}{w^U} = 3$	$\rho = 0.05$	$\rho = 0.08$	$\frac{w^S}{w^U} = 2$	$\rho = 0.05$	$\rho = 0.08$	$\frac{w^S}{w^U} = 4$	$\rho = 0.05$	$\rho = 0.08$
$r(0) = 0.10$	1.61	1.77	1.25	1.32		1.97	2.23	
$r(0) = 0.15$	1.46	1.60	1.18	1.24		1.74	1.97	
$r(0) = 0.20$	1.37	1.49	1.14	1.19		1.60	1.80	

32 Inferred Productivity Difference Estimates, fragile (A-B) case

$\frac{w^S}{w^U} = 3$	$\rho = 0.05$	$\rho = 0.08$	$\frac{w^S}{w^U} = 2$	$\rho = 0.05$	$\rho = 0.08$	$\frac{w^S}{w^U} = 4$	$\rho = 0.05$	$\rho = 0.08$
$r(0) = 0.10$	1.79	2.16	1.38	1.57		2.31	2.78	
$r(0) = 0.15$	1.59	1.87	1.26	1.41		1.92	2.34	
$r(0) = 0.20$	1.47	1.70	1.20	1.32		1.71	2.06	

37
 38 The two tables report implied cross country differences in measured TFP for the parameter
 39 values we used to calculate the cross-country differences in output-per-capita. Recall that the
 40 values are calculated for countries with identical technology sets, vanishingly small differences
 41 in fundamentals, and under the assumption that the researcher perfectly measures the stock
 42 of managerial skills and their returns at a point in time. These tables illustrate how the
 43 accumulation of managerial skill could generate perceived differences in TFP between countries
 44 with similar fundamentals, but which have diverged in their development path. These numbers
 45

46 27. The fraction of income which is a return to managerial skill is calculated as discussed in the previous
 47 subsection, that is, it is calculated as $(\frac{s_{it} F_1(s_{it}, 1-s_{it})}{F(s_{it}, 1-s_{it})})(1 - \frac{F_2(s_{it}, 1-s_{it})}{F_1(s_{it}, 1-s_{it})})$ where s is the fraction of managers in the
 48 population.

1 are also a direct indication of the bias in growth accounting, as the data generated here are for
2 countries with identical technology sets.

3 While these differences are significant, they are still far from explaining the enormous
4 differences in imputed TFP across all countries of the world, for example the 8 fold differences
5 between the lowest and highest productivity countries reported in Hall and Jones (1999).
6 They do indicate, however, that the purposeful accumulation of on-the-job skills could still be
7 significant. Even this conservative estimation could easily explain a quarter of the reported
8 productivity differential between the highest and lowest output countries. An idea of the
9 effective magnitude explained can be obtained by considering groups of countries with
10 productivity around one half of that in the most productive countries. In Hall and Jones
11 estimates, this corresponds to explaining the productivity differential between the top group of
12 countries, including the US, and South American countries.

13 14 5. CONCLUSION 15

16 We have presented a model of managerial skill acquisition where unskilled individuals invest
17 in skills by accepting employment at firms where managerial skills are present. This process
18 of skills acquisition is related to, but distinct from, a previously large literature on skills
19 acquisition through learning-by-doing. Learning-by-doing emphasizes increases in productivity
20 arising from repetition of tasks through time. Our process of managerial skills acquisition
21 involves learning from working with a manager, and thus arises when undertaking distinct
22 and complementary tasks. This process may have potentially important implications for
23 understanding why technologies may be so difficult to transplant across countries. Unlike
24 the learning-by-doing literature, and previous studies which have similarly emphasized the role
25 of embodied skills (as opposed to disembodied knowledge), the process we posit here yields
26 implications that appear to be more consistent with the empirical record. Specifically, it implies
27 there are weak incentives for skilled migration from skill abundant countries, there are strong
28 incentives for skilled migration from skill poor countries (brain-drain), that we should expect to
29 observe low manager/worker ratios in managerial skill rich countries, and that the relationship
30 between skill premia and skill abundance when new technologies are introduced should be
31 positive. A key factor in explaining all of these effects, which is novel in our formulation of
32 learning, is the presence of a teaching premium for managers in rich countries, which will
33 generally be higher than in poor ones. This raises managerial productivity in countries where
34 skills are actively sought, and where skills are high.

35 Although this process of managerial skill acquisition is modeled in a manner that allows
36 knowledge to be entirely internalized by individuals, it nevertheless leads to outcomes
37 resembling those obtained in endogenous growth models with market imperfections. For
38 example, agents will purposefully accumulate managerial skill knowing full well that the
39 return to these skills will dissipate over time and, in the steady state, possibly equal zero.
40 This process also has the potential to lead astray growth accounting exercises since differences
41 across countries in managerial skill would most likely be erroneously attributed to differences
42 in total factor productivity. We have shown how growth accounting should be modified
43 in order to capture this process, and provided a quantitative estimate of the magnitude of
44 cross country productivity differences that can be explained by such a process of on-the-
45 job skills acquisition. The calibration exercise makes clear why returns to experience, which
46 had previously been thought to be a good proxy for the value of skills learned on the
47 job, will not generally capture the full contribution of such skills to cross-country income
48 differences.

APPENDIX A. DYNAMICS OF λ AND S

A.1. The Dynamics of λ

The $\dot{\lambda} = 0$ can take on two configurations. In the first case, this line lies only in the region where $\beta = 1$ and is given by $\dot{\lambda} = F_2(S, (1-S)) - F_1(S, (1-S)) + \lambda \left[\frac{(\delta + \rho) + \Omega(S, (1-S))}{\{\Omega_2(S, (1-S)) - \Omega_1(S, (1-S))\}} (1-S) \right]$. In this case, the $\dot{\lambda} = 0$ line is a downward sloping line asymptoting at $S = 0$ and meeting the S axis at $S = S'$, as depicted in Figure 9. Beyond the S' locus, $\gamma < 1$, and λ must take value 0. The directional arrows indicate the movements of λ .

Crossing points are defined by the intersection of the $\dot{\lambda} = 0$ equation with the $\lambda_{\beta=1}$ equation. These occur at values of S defined by:

$$\frac{\bar{w} - F_2(S, (1-S))}{\Omega + \Omega_2(S, (1-S))(1-S)} = \frac{F_1(S, (1-S)) - F_2(S, (1-S))}{\delta + \rho + \Omega(S, (1-S)) + [\Omega_2(S, (1-S)) - \Omega_1(S, (1-S))](1-S)}. \tag{A1}$$

There may be multiple roots to this equation.

In the $\beta < 1$ region, the value of β is implicitly defined by the condition $\bar{w} - F_2(S_t, \beta_t(1-S_t)) = \lambda_t [\Omega(S_t, \beta_t(1-S_t)) + \Omega_2(S_t, \beta_t(1-S_t))\beta_t(1-S_t)]$, and we have $\dot{\lambda} = (\rho + \delta)\lambda + \bar{w} - F_1(S, \beta(1-S)) - \lambda\beta(1-S)(\Omega_1(S, \beta(1-S)))$. The $\dot{\lambda} = 0$ locus is defined for $\beta < 1$ only if there exist points of intersection defined by equation (A1). Denote the set of all S at such intersection points by Δ . For values of S, λ below the locus $\lambda_{\beta=1}$ the $\dot{\lambda} = 0$ line is given by:

$$\lambda = \frac{F_1(S, \beta(1-S)) - \bar{w}}{\rho + \delta - \Omega_1(S, \beta(1-S))\beta(1-S)}. \tag{A2}$$

It follows from homotheticity that such locuses are horizontal lines

Lemma A1. For $\beta < 1, \dot{\lambda} = 0$ implies $\frac{d\lambda}{dS} = 0$.

A situation with two crossing points is depicted in Figure 10. Once again, the arrows depict the dynamics of λ .

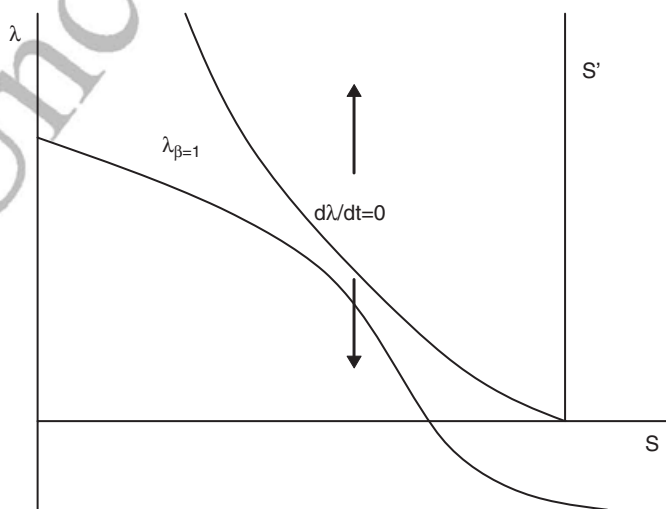


FIGURE 9

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

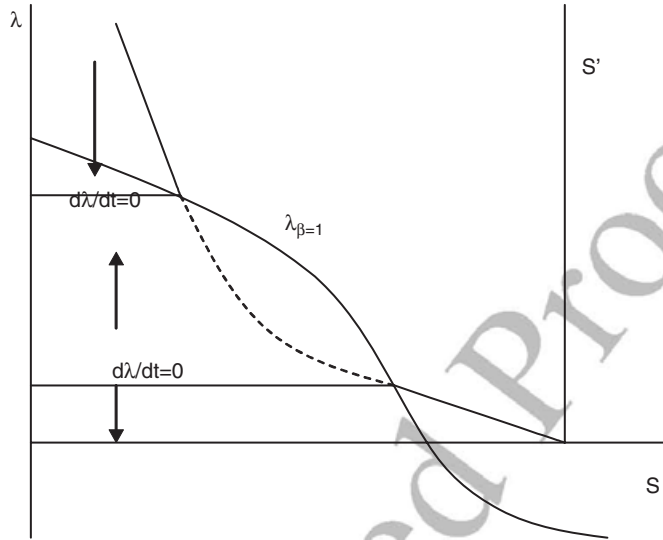


FIGURE 10

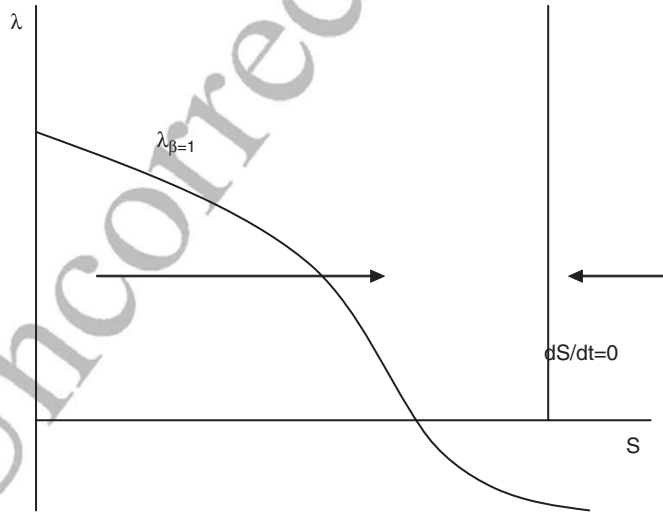


FIGURE 11

A.2. Dynamics of S

The $\dot{S} = 0$ also has one of two different configurations that depend on whether it crosses the $\lambda_{\beta=1}$ line. In the case where $\dot{S} = 0$ falls entirely in the region where $\beta = 1$, then the $\dot{S} = 0$ equation is given by the unique solution of

$$\Omega(S, (1 - S))(1 - S) = \delta S, \tag{A3}$$

denoted $\bar{S} \in (0, 1)$. It follows that the locus is bounded away from 1 because $\Omega(S, 0)(0) = 0$. In this case, the $\dot{S} = 0$ equation corresponds to a vertical line at \bar{S} , as depicted in Figure 11. Note that \bar{S} can be either to the right or left of S' .

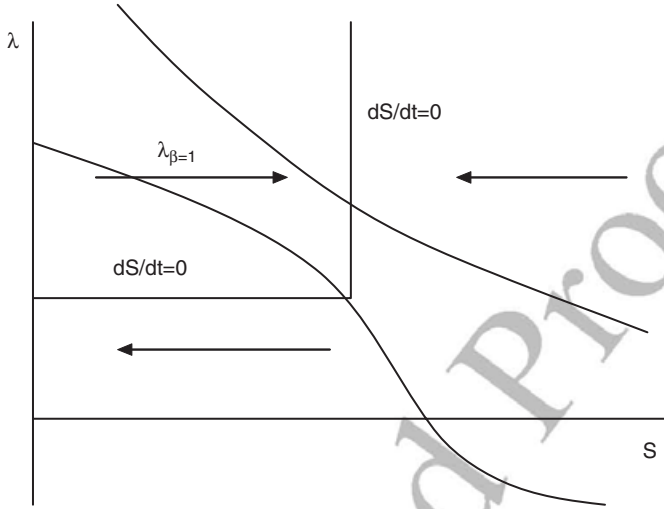


FIGURE 12

It is also possible for there to be an intersection between $\dot{S} = 0$ and $\lambda_{\beta=1}$, which is defined by the pair $S = \bar{S}$ and $\lambda = \frac{\bar{w} - F_2(\bar{S}, (1-\bar{S}))}{\Omega + \Omega_2(\bar{S}, (1-\bar{S}))}$. When such a crossing does not exist in the positive orthant, i.e. when $\bar{w} - \theta F_2(\bar{S}, 1 - \bar{S}) < 0$, the movement of the system is always towards $S = \bar{S}$ as depicted in Figure 11. Conversely, if a crossing exists, it is unique, and then the following lemma shows that the $\dot{S} = 0$ line is a horizontal schedule in the region where $\beta < 1$.

Lemma A2. For $\beta < 1$, $\dot{S} = 0$ implies $\frac{d\lambda}{dS} = 0$.

Figure 12 depicts the dynamics for S when such a point of intersection exists.

The following two lemmas establish that the $\lambda_{\beta=1}$ locus, and that the $\lambda = 0$ in the region where $\beta = 1$ are both downward sloping.

Lemma A3. $\frac{d\lambda_{\beta=1}}{dS} < 0$

Lemma A4. $\frac{d\lambda}{dS} (\lambda = 0|_{\beta=1}) < 0$

APPENDIX B. PROOFS

Proof of Lemma A1. Rewrite expression (A2) as $\lambda = \frac{F_1(S, \beta(1-S)) - \bar{w}}{\rho + \delta - \Omega_1(S, \beta(1-S))}$ and differentiate with respect to S yielding:

$$\frac{d\lambda}{dS} = \left(F_{11} + \left(\frac{d\beta}{dS} (1-S) - \beta \right) F_{12} \right) (\rho + \delta + \tilde{\Omega}_1) + (F_1 - \bar{w}) \left(\tilde{\Omega}_{11} + \tilde{\Omega}_{12} \left(\frac{d\beta}{dS} (1-S) - \beta \right) \right). \quad (B1)$$

Differentiating the first-order condition yields $\left(\frac{d\beta}{dS} (1-S) - \beta \right) = -\frac{F_{21} + \frac{d\lambda}{dS} \tilde{\Omega}_2 + \lambda \tilde{\Omega}_{21}}{F_{22} + \tilde{\Omega}_{22} \lambda}$. Substituting into (B1) and rearranging yields:

$$\begin{aligned} \frac{d\lambda}{dS} \left(1 + \frac{\tilde{\Omega}_2}{F_{22} + \tilde{\Omega}_{22} \lambda} (F_{12} + \tilde{\Omega}_{12}) \right) &= \left(F_{11} - \frac{F_{21} + \lambda \tilde{\Omega}_{21}}{F_{22} + \lambda \tilde{\Omega}_{22}} F_{12} \right) (\rho + \delta + \tilde{\Omega}_1) \\ &+ (F_1 - \bar{w}) \left(\tilde{\Omega}_{11} - \frac{F_{21} + \lambda \tilde{\Omega}_{21}}{F_{22} + \lambda \tilde{\Omega}_{22}} \tilde{\Omega}_{12} \right). \end{aligned} \quad (B2)$$

1 Since $\tilde{\Omega}$ and F are HD1, we have $F_{11}S + F_{12}\beta(1-S) = F_{21}S + F_{22}\beta(1-S) = \tilde{\Omega}_{11}S + \tilde{\Omega}_{12}\beta(1-S) = \tilde{\Omega}_{21}S + \tilde{\Omega}_{22}\beta(1-S) = 0$. These equalities imply that the terms in large parentheses on the right-hand side of (B2) equal zero, so that $\frac{d\lambda}{dS} = 0$. \parallel

2
3
4
5 *Proof of Lemma A2.* $\dot{S} = 0$ implies $\Omega(S, \beta(1-S))\beta(1-S) \equiv \tilde{\Omega}(S, \beta(1-S)) = \delta S$. Differentiating yields
6 $\tilde{\Omega}_1 + \tilde{\Omega}_2 \left(\frac{d\beta}{dS} (1-S) - \beta \right) = \delta$. As in the previous Lemma, differentiating the first order condition yields
7 $\left(\frac{d\beta}{dS} (1-S) - \beta \right) = -\frac{F_{21} + \frac{d\lambda}{dS} \tilde{\Omega}_2 + \lambda \tilde{\Omega}_{21}}{F_{22} + \tilde{\Omega}_{22}\lambda}$. Combining these two expressions leads to:

$$8 \quad \text{sign} \left(\frac{d\lambda}{dS} \right) = \text{sign} \left((\delta - \tilde{\Omega}_1) (F_{22} + \tilde{\Omega}_{22}\lambda) + (F_{21} + \lambda \tilde{\Omega}_{21}) \tilde{\Omega}_2 \right). \quad (B3)$$

9
10
11 Since $\tilde{\Omega}$ is HD1, $\tilde{\Omega}_1 S + \tilde{\Omega}_2 \beta(1-S) = \tilde{\Omega} = \delta S$, where the last inequality holds at $\dot{S} = 0$. Using this expression to substitute for $\tilde{\Omega}_1$ in (B3) yields:

$$12 \quad \text{sign} \left(\frac{d\lambda}{dS} \right) = \text{sign} \left(\left(\tilde{\Omega}_2 \beta \frac{1-S}{S} \right) (F_{22} + \tilde{\Omega}_{22}\lambda) + (F_{21} + \lambda \tilde{\Omega}_{21}) \tilde{\Omega}_2 \right). \quad (B4)$$

13
14
15
16 Since F_i and $\tilde{\Omega}_i$ are HD0, we have $SF_{21} + \beta(1-S)F_{22} = S\tilde{\Omega}_{21} + \beta(1-S)\tilde{\Omega}_{22} = 0$. Using these to substitute into
17 (B4) yields $\text{sign} \left(\frac{d\lambda}{dS} \right) = \text{sign} \left((\tilde{\Omega}_2 \beta \frac{1-S}{S}) (F_{22} + \tilde{\Omega}_{22}\lambda) - (\beta \frac{1-S}{S} F_{22} + \lambda \beta \frac{1-S}{S} \tilde{\Omega}_{22}) \tilde{\Omega}_2 \right) = 0$. \parallel

AQ11

18
19
20 *Proof of Lemma A3.* $\frac{d\lambda_{\beta=1}}{dS} = \text{sign}[-(\Omega - \Omega_2(1-S)) (F_{21} - F_{22}) - (\bar{w} - F_2) (\Omega_1 - \Omega_2 + (\Omega_{21} - \Omega_{22})(1-S) - \Omega_2)]$. The first term in the square brackets is negative. From HDO of Ω we have:

$$21 \quad (\Omega_{21} - \Omega_{22})(1-S) + (\Omega_1 - \Omega_2) = (\Omega_{12} - \Omega_{11})S \quad (B5)$$

22
23 which implies that the second term becomes

$$24 \quad -(-\Omega_2 + (\Omega_{12} - \Omega_{11})S). \quad (B6)$$

25
26 From the homogeneity of degree -1 of the function Ω_1 we have $\Omega_{11}S + \Omega_{12}(1-S) = -\Omega_1$. Substituting from (B5)
27 we obtain $\Omega_{11} \frac{S^2}{1-S} + S\Omega_{12} = \Omega_2$, implying that $S\Omega_{12} - \Omega_2 > 0$. Using this and the fact that $\Omega_{11} < 0$, the second
28 term, expression (B6), is negative so that the whole expression is negative. \parallel

29
30
31 *Proof of Lemma A4.* $\lambda = 0|_{\beta=1} \Rightarrow \lambda = \frac{F_1 - F_2}{\delta + \rho + \Omega_2 - \Omega_1}$ and $(\text{sign}) \frac{d\lambda}{dS} = \text{sign} \left[\frac{(F_{11} - F_{12} - F_{21} + F_{22})(\delta + \rho + \tilde{\Omega}_2 - \tilde{\Omega}_1)}{-(F_1 - F_2)(\tilde{\Omega}_{21} - \tilde{\Omega}_{22} - \tilde{\Omega}_{12} + \tilde{\Omega}_{12})} \right]$.
32 Since $\tilde{\Omega}_{ij} > 0$, $\tilde{\Omega}_{ii} < 0$, $F_{ii} < 0$ and $F_{ij} > 0$ it follows that $(\text{sign}) \frac{d\lambda}{dS} < 0$. \parallel

33
34
35 *Proof of Proposition 1.* We first show that it is not possible for the $\dot{\lambda} = 0$ to both intersect $\lambda_{\beta=1}$ and also have
36 an intersection at $\dot{S} = 0$ in the $\beta = 1$ region at an S value below the smallest element of Λ . That is we first rule out
37 the case depicted in Figure 13 below.

38 For this proof it is simpler to use the function $\tilde{\Omega}(S, 1-S) \equiv \Omega(S, 1-S)(1-S)$. Since Ω is HDO the function
39 $\tilde{\Omega}$ is HD1. Assign the following λ values: For \bar{S} defined in (A3) we have $\lambda_4 = \frac{F_1 - F_2}{\delta + \rho + \Omega_2 - \Omega_1}$ and $\lambda_3 = \frac{\bar{w} - F_2}{\Omega_2}$. Consider
40 any two elements S_1, S_2 of Λ : and consider an S^* with $S_1 < S^* < S_2$, such that the $\lambda = 0$ function defined for
41 $\beta = 1$ lies below the $\lambda_{\beta=1}$ line. Then define $\lambda_2 = \frac{\bar{w} - F_2^*}{\tilde{\Omega}_2^*}$ and $\lambda_1 = \frac{F_1^* - F_2^*}{\delta + \rho + \tilde{\Omega}_2^* - \tilde{\Omega}_1^*}$, where $F^* \equiv F(S^*, 1-S^*)$ and
42 $\tilde{\Omega}^* \equiv \tilde{\Omega}(S^*, 1-S^*)$. Clearly, for this case to exist it is necessary that: $\lambda_4 > \lambda_3 > \lambda_2 > \lambda_1$.

43 From the HD1 of $\tilde{\Omega}$ we have $\tilde{\Omega}_1 S + \tilde{\Omega}_2 (1-S) = \tilde{\Omega}$. At $S = \bar{S}$, $\tilde{\Omega} = \delta \bar{S}$, combining these implies that, at $S = \bar{S}$

$$44 \quad \tilde{\Omega}_2 = (\delta - \tilde{\Omega}_1) \frac{\bar{S}}{1 - \bar{S}}. \quad (B7)$$

45 For the remainder of the proof $\tilde{\Omega}$ and F will denote the respective functions evaluated at $S = \bar{S}$. Applying these,
46 $\lambda_4 > \lambda_3$ implies that $\frac{F_1 - F_2}{\delta + \rho + \Omega_2 - \Omega_1} > \frac{\bar{w} - F_2}{\Omega_2} \Rightarrow \frac{F_1 - F_2}{\delta + \rho + (\delta - \tilde{\Omega}_1) \frac{\bar{S}}{1 - \bar{S}} - \tilde{\Omega}_1} > \frac{\bar{w} - F_2}{(\delta - \tilde{\Omega}_1) \frac{\bar{S}}{1 - \bar{S}}} \Rightarrow \frac{F_1 - F_2}{\rho + (\delta - \tilde{\Omega}_1) \frac{1}{1 - \bar{S}}} > \frac{\bar{w} - F_2}{(\delta - \tilde{\Omega}_1) \frac{\bar{S}}{1 - \bar{S}}} \Rightarrow$
47 $\frac{F_1 - F_2}{(1 - \bar{S})\rho + (\delta - \tilde{\Omega}_1)} > \frac{\bar{w} - F_2}{(\delta - \tilde{\Omega}_1)\bar{S}} \Rightarrow (\delta - \tilde{\Omega}_1)\bar{S}F_1 - (\delta - \tilde{\Omega}_1)\bar{S}F_2 > (1 - \bar{S})\rho(\bar{w} - F_2) + (\delta - \tilde{\Omega}_1)(\bar{w} - F_2) \Rightarrow (\delta - \tilde{\Omega}_1)$

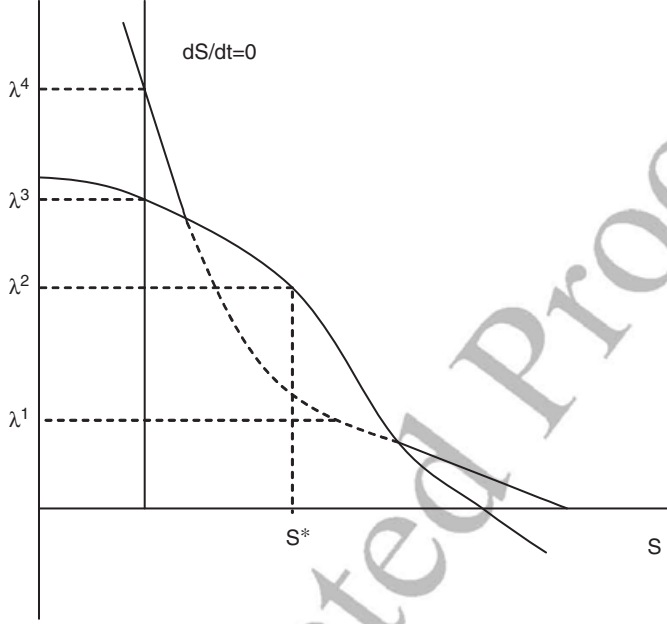


FIGURE 13

$$\begin{aligned} \bar{S}F_1 - (\delta - \tilde{\Omega}_1)\bar{S}F_2 &> (1 - \bar{S})\rho(\bar{w} - F_2) + \bar{S}(\delta - \tilde{\Omega}_1)(\bar{w} - F_2) + (1 - \bar{S})(\delta - \tilde{\Omega}_1)(\bar{w} - F_2) \Rightarrow (\delta - \tilde{\Omega}_1) \\ \bar{S}(F_1 - \bar{w}) &> (1 - \bar{S})(\delta - \tilde{\Omega}_1 + \rho)(\bar{w} - F_2) \Rightarrow \frac{F_1 - \bar{w}}{\delta - \tilde{\Omega}_1 + \rho} > \frac{\bar{w} - F_2}{(\delta - \tilde{\Omega}_1)\bar{S}} \text{ which implies:} \end{aligned}$$

$$\Rightarrow \frac{F_1 - \bar{w}}{\delta - \tilde{\Omega}_1 + \rho} > \frac{\bar{w} - F_2}{\tilde{\Omega}_2}. \quad (B8)$$

Now notice that

$$\frac{F_1^* - F_2^*}{\delta + \rho + \tilde{\Omega}_2^* - \tilde{\Omega}_1^*} > \frac{F_1^* - F_2^*}{\delta + \rho - \tilde{\Omega}_1^*} > \frac{F_1^* - \bar{w}}{\delta + \rho - \tilde{\Omega}_1^*} > \frac{F_1^* - \bar{w}}{\delta + \rho - \tilde{\Omega}_1(\bar{S})}. \quad (B9)$$

The first inequality follows since $\tilde{\Omega}_2^* < 0$, the second since $\bar{w} > F_2^*$ as $\beta < 1$ for $S = S^*$, the third since $\tilde{\Omega}_1^* < \tilde{\Omega}_1(\bar{S})$ as $S^* > \bar{S}$. Notice also that:

$$\frac{\bar{w} - F_2^*}{\tilde{\Omega}_2^*} < \frac{\bar{w} - F_2^*}{\tilde{\Omega}_2(\bar{S})}, \quad (B10)$$

since $\tilde{\Omega}_2^* > \tilde{\Omega}_2(\bar{S})$. Using (B9) and (B10) it then follows that $\lambda_2 > \lambda_1$ implies:

$$\frac{\bar{w} - F_2^*}{\tilde{\Omega}_2(\bar{S})} > \frac{F_1^* - \bar{w}}{\delta + \rho - \tilde{\Omega}_1(\bar{S})}. \quad (B11)$$

Once again utilizing (B7) inequalities (B8) and (B11) can be expressed respectively as:

$$\frac{(F_1 - \bar{w})\bar{S}}{\delta - \tilde{\Omega}_1 + \rho} > \frac{(\bar{w} - F_2)(1 - \bar{S})}{\delta - \tilde{\Omega}_1} \quad (B12)$$

$$\frac{(\bar{w} - F_2^*)(1 - \bar{S})}{\delta - \tilde{\Omega}_1} > \frac{(F_1^* - \bar{w})\bar{S}}{\delta + \rho - \tilde{\Omega}_1(\bar{S})}. \quad (B13)$$

Adding the two inequalities above makes the \bar{w} terms cancel and yields $\frac{F_1 \bar{S}}{\delta - \Omega_1 + \rho} - \frac{F_2^*(1 - \bar{S})}{\delta - \Omega_1} > \frac{\bar{S} F_1^*}{\delta + \rho - \Omega_1(\bar{S})} - \frac{F_2(1 - \bar{S})}{\delta - \Omega_1}$. Rearranging terms in this expression implies $F_1 \bar{S} (\delta - \tilde{\Omega}_1) - F_2^* (1 - \bar{S}) (\delta - \tilde{\Omega}_1 + \rho) > \bar{S} F_1^* (\delta - \tilde{\Omega}_1) - F_2 (1 - \bar{S}) (\delta + \rho - \tilde{\Omega}_1(\bar{S}))$ which implies $F_1 \bar{S} + F_2 (1 - \bar{S}) > \bar{S} F_1^* + F_2^* (1 - \bar{S})$. But since $S^* > \bar{S}$ this is false by the concavity of F .

With this case ruled out we then have the following exhaustive set of possibilities:

Either (1) $\dot{\lambda} = 0$ does not intersect $\lambda_{\beta=1}$. In that case, it follows from Lemma 1, that the downward sloping $\dot{\lambda} = 0$ locus intersects the vertical $\dot{S} = 0$ in the $\beta = 1$ region only once (Figure 4). This is a unique steady state in which all individuals are employed in the modern technology.

Or (2) $\dot{\lambda} = 0$ intersects $\lambda_{\beta=1}$ and does not intersect $\dot{S} = 0$, in the $\beta = 1$ region. In this case, ignoring the parameter singularity where the vertical locus $\dot{S} = 0$ and $\dot{\lambda} = 0$ overlap, there is no intersection between $\dot{S} = 0$ and $\dot{\lambda} = 0$ in the region $\beta < 1$ either. However at $S = 0$, $\bar{w} - F_2(0, 0) = \bar{w} > \lambda_t \Omega$, so that $\beta_t = 0$. But if $\beta_t = 0$ then $\dot{S} = 0$ at $S = 0$, since S is constrained to be non-negative. Consequently a unique steady state exists at the intersection of the $\dot{\lambda} = 0$ locus and the vertical axis, as depicted in Figure 5.

A final possibility is situation (3) that $\dot{\lambda} = 0$ intersects $\lambda_{\beta=1}$ and intersects $\dot{S} = 0$ at a value of S to the right of the last intersection point between $\lambda_{\beta=1}$ and $\dot{\lambda} = 0$. An intersection at any other point has been ruled out by the first part of the proof above. At $S = 0$, it is again the case that $\bar{w} > \lambda_t \Omega$, so that $\dot{S} = 0$ since S is constrained to be non-negative. A steady state thus exists at the intersection of the $\dot{\lambda} = 0$ locus and the vertical axis, as depicted in Figure 6, point B. But another steady state exists in the $\beta = 1$ region at the intersection of the two loci. \parallel

Proof of Proposition 2. Proof of part a: The $\dot{\lambda} = 0$ crosses the $\lambda_{\beta=1}$ locus if and only if there exists at least one value of S for which,

$$\frac{\bar{w} - F_2(\cdot)}{\Omega(\cdot) + \Omega_2(\cdot)(1 - S)} - \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + [\Omega_2(\cdot) - \Omega_1(\cdot)](1 - S)} > 0, \quad (\text{B14})$$

i.e. if the set Λ is non-empty. Suppose that $\rho + \delta = 0$. Then, necessarily, condition (B14) fails. Because $F_1 > \bar{w}$ and $\Omega_1 > 0$. Since the second expression is monotonically decreasing in $\rho + \delta$ and the first is unaffected, there necessarily exists a unique value of $\rho + \delta$ such that

$$\frac{\bar{w} - F_2(\cdot)}{\Omega(\cdot) + \Omega_2(\cdot)(1 - S)} - \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + [\Omega_2(\cdot) - \Omega_1(\cdot)](1 - S)} = 0. \quad (\text{B15})$$

Let k^* denote the value of $\rho + \delta$ solving (B15) with equality. If and only if $\rho + \delta < k^*$ the functions do not cross and the situation in Figure 4 occurs. In these, the system's unique steady state, involves no one working in the traditional technology.

Proof of part b: S' is defined from equation (1). Consider \bar{S} defined in (A3). If and only if $\bar{S} > S'$, then the $\dot{S} = 0$ locus lies to the right of the S' locus. This configuration is not depicted but similar to that in Figure 4 with the relative positions of the $\dot{S} = 0$ and $S = S'$ lines reversed. In this steady state, $\lambda = 0$. In any configuration where $\dot{S} = 0$ lies to the left of S' , the steady state involves $\lambda > 0$. Thus $\delta^* = \frac{\Omega(S', 1 - S')}{S'}$.

Proof of part c: δ^{**} solves $F_2(\bar{S}, 1 - \bar{S}) = \bar{w}$ and $\delta^{**} = \frac{\Omega(\bar{S}, 1 - \bar{S})}{\bar{S}}$. For $\delta < \delta^{**}$, the $\dot{S} = 0$ locus does not intersect the $\lambda_{\beta=1}$ line and an interior steady state exists. If $\delta + \rho < k^*$, there is also no intersection with the $\dot{\lambda} = 0$ locus and the unique steady state is in the interior, as in Figure 4. If $\delta + \rho \geq k^*$, there is an intersection and the case of Figure 6 applies, implying hysteresis. For $\delta \geq \delta^{**}$, the $\dot{S} = 0$ locus intersects the $\lambda_{\beta=1}$ line. Note that, from condition (2) $\delta^{**} > \delta^*$. Whether the economy is in the hysteresis case, or the stable traditional steady state with full use of traditional technology depends on the location of the $\dot{\lambda} = 0$ line. The critical value of ρ , denoted ρ^* solves $\frac{\bar{w} - F_2(\bar{S}, 1 - \bar{S})}{\Omega(\bar{S}, 1 - \bar{S}) + \Omega_2(\bar{S}, 1 - \bar{S})(1 - \bar{S})} = \frac{F_1(\bar{S}, 1 - \bar{S}) - F_2(\bar{S}, 1 - \bar{S})}{\delta + \rho + \Omega(\bar{S}, 1 - \bar{S}) + [\Omega_2(\bar{S}, 1 - \bar{S}) - \Omega_1(\bar{S}, 1 - \bar{S})](1 - \bar{S})}$, which implies that $\rho^* = \frac{F_1(\cdot)[\Omega(\cdot) + \Omega_2(\cdot)(1 - \bar{S})] + F_2(\cdot)[\delta - \Omega_1(1 - \bar{S})] - \bar{w}\Omega(\cdot) + (\Omega_2(\cdot) - \Omega_1(\cdot))(1 - \bar{S}) + \delta}{\bar{w} - F_2(\cdot)}$. Note that since $\alpha F_1 + (1 - \alpha) F_2 > \bar{w}$ for any $\alpha \in [0, 1]$, $\rho^* > 0$. For $\rho > \rho^*$, the $\dot{\lambda} = 0$ locus intersects the $\lambda_{\beta=1}$ line below the intersection with $\dot{S} = 0$, so that the situation in Figure 5 occurs, and the unique steady state is full use of the traditional technology. Under the converse, the situation in Figure 5 occurs and hysteresis is the outcome. \parallel

Proof of Proposition 3. Skilled managers: We prove these individuals do not migrate from the rich country by demonstrating that the steady-state skilled wage, w^s , is strictly decreasing in ρ , ceteris paribus. From Proposition 3 recall that the steady-state allocation of labour varies with the sum $\rho + \delta$ depicted by the downward sloping straight line in Figure 7. For given $\delta < \delta'$, there thus exists a unique value of ρ , denote it $\rho^c \equiv k^* - \delta$, such that for $\rho > \rho^c$,

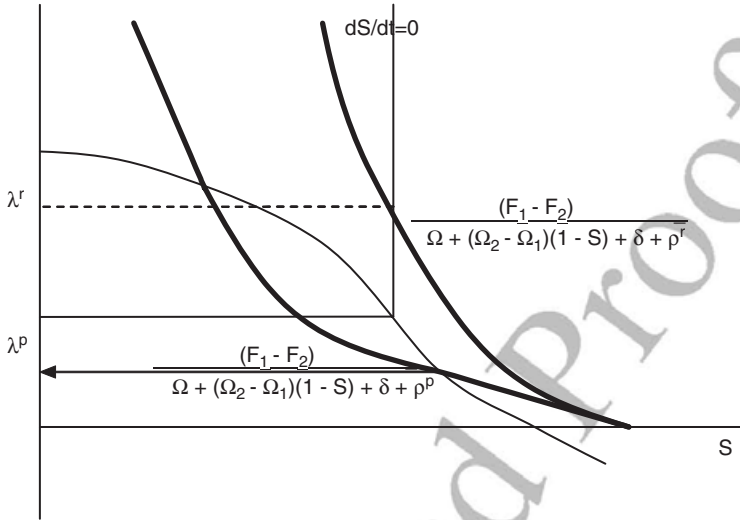


FIGURE 14

$\beta = 0$ (corresponding to the configuration in Figure 5) and the country is poor. For $\rho \leq \rho^c$, $\beta = 1$ (corresponding to the configurations in Figures 3 or 5) and the country is rich. For $\rho \leq \rho^c$ then $w^s(\rho) = F_1(\cdot) + \lambda\Omega_1(\cdot)(1 - S)$. Where steady-state S is determined by $\delta S = \Omega(\cdot)(1 - S)$ and $\lambda = \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + \Omega_2(\cdot)(1 - S) - \Omega_1(\cdot)(1 - S)}$. This implies: $w^s(\rho) = F_1(\cdot) + \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + \Omega_2(\cdot)(1 - S) - \Omega_1(\cdot)(1 - S)}\Omega_1(\cdot)(1 - S)$. Since steady-state S is unaffected by ρ it is immediate from the above that $\frac{dw^s(\rho)}{d\rho} < 0$.

For $\rho > \rho^c$,

$$w^s(\rho) = F_1(Z, 1) + \lambda\tilde{\Omega}_1(Z, 1), \tag{B16}$$

where $Z \equiv \frac{\beta S}{1 - S}$. The value of λ is then given by $\lambda = \frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)}$, and Z is the smallest root of $\frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \Omega_1(Z, 1)} = \frac{\bar{w} - F_2(Z, 1)}{\Omega_2(Z, 1)}$, corresponding to the value of Z at the point A in Figure 5, or the point λ^p in Figure 14 below. By inspection of the figure it is immediate that, as ρ increases, the curved schedule shifts downwards to the left, implying that λ falls. This implies that Z increases since λ also satisfies $\lambda = \frac{\bar{w} - F_2(Z, 1)}{\Omega_2(Z, 1)}$. Since Z increases and λ falls, it is immediate from (B16) that $w^s(\rho)$ is decreasing in ρ for this range of ρ as well. Consequently, since $w^s(\rho)$ is everywhere strictly decreasing in ρ , there is a strict incentive for skilled individuals to migrate from the poor to the rich country.

Unskilled workers: The wage to working in the traditional technology is the same in each country, \bar{w} . In the poor country, all individuals strictly prefer to work with the traditional technology, $\beta = 0$. In the rich country, the unskilled strictly prefer to work in the modern technology instead of working with the traditional technology which is still available to them, i.e. $\beta = 1$. Consequently, returns are higher to the unskilled in the rich country.

Proof of Corollary 1. Increasing F has identical effects to decreasing ρ , it shifts out the $\dot{\lambda} = 0$ schedule, shifts in the $\lambda_{\beta=1}$ schedule and leaves $\dot{S} = 0$ unchanged, so the analysis of Proposition 3 applies. There are thus three possible cases to consider: (1) A country previously not using the modern technology under F does so under \hat{F} ; (2) A country not using the modern technology under F does not change to using it under \hat{F} ; and (3) A country already using the modern technology under F continues to do so under \hat{F} . Case (1). The skill premium strictly rises when the economy transitions into using \hat{F} . Skill supply also strictly increases. The premium eventually falls as the economy converges to the new steady state, but it remains strictly above the skill premium under F . Case (2) If already using the new technology, i.e. a situation corresponding to Figure 4, then the $\dot{\lambda} = 0$ locus intersects with $\dot{S} = 0$ at a higher λ so the skill premium is strictly higher, and skill supply is unchanged. Case (3) if an economy remains in the traditional technology under \hat{F} since $\dot{\lambda} = 0$ rises, the shadow value of skills strictly increases and skill levels are unchanged.

Proof of Proposition 4. At steady-state A all individuals work in the modern sector therefore $w^s(A) = F_1(S^A, 1 - S^A) + \lambda \Omega_1(S^A, 1 - S^A)(1 - S^A)$. Where steady-state S is determined by $\delta S^A = \Omega(S^A, 1 - S^A)(1 - S^A)$ and $\lambda = \frac{F_1((S^A, 1 - S^A)) - F_2((S^A, 1 - S^A))}{\delta + \rho + \Omega((S^A, 1 - S^A)) + [\Omega_2((S^A, 1 - S^A)) - \Omega_1((S^A, 1 - S^A))](1 - S^A)} \equiv \lambda^A$. This implies:

$$w^s(A) \equiv F_1((S^A, 1 - S^A)) + \lambda^A \tilde{\Omega}_1((S^A, 1 - S^A)). \quad (B17)$$

At the point B , the introduction of a small number of skilled individuals implies:

$$w^s(B) = F_1(Z, 1) + \lambda^B \tilde{\Omega}_1(Z, 1), \quad (B18)$$

where $Z \equiv \frac{\beta S}{1 - S}$. The value of λ is then given by $\lambda = \frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \tilde{\Omega}_1(Z, 1)}$, and Z is the smallest root of $\frac{F_1(Z, 1) - \bar{w}}{\delta + \rho - \tilde{\Omega}_1(Z, 1)} = \frac{\bar{w} - F_2(Z, 1)}{\tilde{\Omega}_2(Z, 1)}$. At S^A since $\beta = 1$ we know that:

$$\bar{w} < F_2((S^A, 1 - S^A)) + \lambda^A \tilde{\Omega}_2((S^A, 1 - S^A)) \quad (B19)$$

After the introduction of a small number of skilled individuals at B , since $\beta < 1$, we know that

$$F_2(Z, 1) + \lambda^B \tilde{\Omega}_2(Z, 1) = \bar{w}. \quad (B20)$$

But since by inspection of Figure 6 it is immediate that $\lambda^B > \lambda^A$ it follows from (B19) and (B20) that $F_2(Z, 1) + \tilde{\Omega}_2(Z, 1) < F_2(S^A, 1 - S^A) + \tilde{\Omega}_2(S^A, 1 - S^A)$. This implies that $Z < \frac{S^A}{1 - S^A}$. It then follows immediately from (B17) and (B18) that $w^s(B) > w^s(A)$. There is a strict incentive for skilled workers to migrate from the rich to the poor country when differences are driven by hysteresis. \parallel

Acknowledgements. We thank Daron Acemoglu, Marios Angeletos, Peter Howitt, Hyeok Jeong, Ashok Kotwal, seminar participants at the UBC macroeconomic workshop, NEUDC (Montreal), MIT, UQAM, ISI Delhi and Brown for helpful comments. We are also extremely grateful to the editor and three anonymous referees for extensive suggestions.

REFERENCES

- ACEMOGLU, D., AGHION, P. and ZILIBOTTI, F. (2004), "Vertical Integration and Distance to Frontier", *Journal of the European Economic Association*, forthcoming.
- ACEMOGLU, D. (2002), "Technical Change, Inequality and the Labor Market", *Journal of Economic Literature*, **40**, 7-72.
- ACEMOGLU, D. and NEWMAN, A. (2002), "The Labor Market and Corporate Structure", *European Economic Review*, **46**, 1733-1756.
- ACEMOGLU, D. and ZILIBOTTI, F. (2001), "Productivity Differences", *Quarterly Journal of Economics*, **115** (3), 563-606.
- AGHION, P. and HOWITT, P. (1992), "A Model of Growth Through Creative Destruction", *Econometrica*, **60**, 323-351.
- ARROW, K. J. (1962a), "Economic Welfare and the Allocation of Re-Sources for Inventions", in R. R. Nelson (ed.) *The Rate and Direction of Inventive Activity* (Princeton University Press and NBER).
- ARROW, K. J. (1962b), "The Economic Implications of Learning by Doing", *Review of Economic Studies*, **29** (3), 155-173.
- AUTOR, D., LEVY, F. and MURNANE, R. (2003), "The Skill Content of Recent Technical Change: An Empirical Exploration", *Quarterly Journal of Economics*, **118**, 1279-1334.
- BEAUDRY, P. and FRANCOIS, P. (2007), "The Economics of Inefficient Technology Use" (NBER Working Paper No. 13500).
- BECKER, G. S. (1975), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 2nd edn (New York: National Bureau of Economic Research and Columbia University Press).
- BENHABIB, J. and SPIEGEL, M. (2003), "Human Capital and Technology Diffusion" (Development Research Institute working paper No. 2003-03, Federal Reserve Bank of San Francisco).
- CHARI, V. V. and HOPENHAYN, H. (1991), "Vintage Human Capital, Growth and the Diffusion of New Technology", *Journal of Political Economy*, **99** (61), 1142-1165.
- EASTERLY, W. (2002), *The Elusive Quest for Growth: Economists Adventures and Misadventures in the Tropics* (Cambridge: MIT press).
- ENOS, M. D., KEHRHAHN, M. T. and BELL, A. (2003), "Informal Learning and the Transfer of Learning: How Managers Develop Proficiency", *Human Resource Development Quarterly*, **14** (4), 369-387.

- 1 FELDMAN, M. (2001), "The Entrepreneurial Event Revisited: Firm Formation in a Regional Context", *Industrial and*
2 *Corporate Change*, **10** (4), 861–891.
- 3 GROSSMAN, G. and HELPMAN, E. (1991), *Innovation and Growth in the Global Economy* (Cambridge: MIT Press).
- 4 HALL, R. and JONES, C. (1999), "Why Do Some Countries Produce So Much More Output per Worker than Others?"
5 *Quarterly Journal of Economics*, **114**, 83–116.
- 6 HECKMAN, J., LOCHNER, L. J. and TODD, P. E., (2008), "Earnings Functions and Rates of Return" (NBER
7 Working Papers No. 13780, National Bureau of Economic Research).
- 8 KELLER, W. (2004), "International Technology Diffusion", *Journal of Economic Literature*, **42** (3), 752–782.
- 9 KERR, W. R. (2008), "Ethnic Scientific Communities and International Technology Diffusion", *The Review of*
10 *Economics and Statistics*, **90** (3), 518–537.
- 11 KLENOW, P. and RODRIGUEZ-CLARE, A. (1997), "The Neoclassical Revival in Growth Economics: Has It Gone
12 Too Far?", in B. S. Bernanke and J. J. Rotemberg (eds.) *NBER Macroeconomics Annual 1997* (Cambridge and
13 London: MIT Press.) 73–103.
- 14 LALL, S. (1999), "Competing with Labour: Skills and Technology in Developing Countries ILO" (Issues in
15 Development Working Paper No. 31).
- 16 LESTER, R. and PIORE, M. (2004), *Innovation: The Missing Dimension* (Cambridge: Harvard University Press).
- 17 LUCAS, R. (1993), "Making a Miracle", *Econometrica*, **61** (2), 251–272.
- 18 MANKIW, N. G. (1995), "The Growth of Nations", *Brookings Papers on Economic Activity*, **1**, 275–310.
- 19 MANKIW, N. G., ROMER, D. and WEIL, D. N. (1992), "A Contribution to the Empirics of Economic Growth",
20 *Quarterly Journal of Economics*, **107** (2), 407–437.
- 21 MCCALL, M., LOMBARDO, M. and MORRISON, A. (1988), *The Lessons of Experience: How Successful Executives*
22 *Develop on the Job* (Lexington: Lexington Press).
- 23 PARENTE, S. and PRESCOTT, E. (2002), *Barriers to Riches* (Cambridge: MIT Press).
- 24 RHEE, Y. and BELOT, T. (1990), "Export Catalysts in Low-Income Countries: A Review of Eleven Success Stories"
25 (World Bank Discussion Paper No. 72).
- 26 ROMER, P. (1995), "The Growth of Nations: Comment", *Brookings Papers on Economic Activity*, **0** (1), 313–320.
- 27 SAXENIAN, A. (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Cambridge:
28 Harvard University Press).
- 29 STOKEY, N. (1988), "Learning by Doing and the Introduction of New Goods", *Journal of Political Economy*, **96** (4),
30 701–717.
- 31 YOUNG, A. (1991), "Learning by Doing and the Dynamic Effects of International Trade", *Quarterly Journal of*
32 *Economics*, **106** (2), 369–405.
- 33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

AQ14

QUERIES TO BE ANSWERED BY AUTHOR

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. DO NOT mark your corrections on this query sheet.

Queries from the Copyeditor:

- AQ1: Author: A running head short title was not supplied; please check if this one is suitable and, if not, please supply a short title of up to X characters that can be used instead.
- AQ2: Author: Please confirm if this abbreviation needs to be spelt out. If yes, please provide the expansion.
- AQ3: Author: The reference “Rosen 1974” has not been listed in the reference list. Please provide the reference details.
- AQ4: Author: Please clarify if the abbreviation “FOCs” should be expanded as “first order conditions” here (because this abbreviation occurs only here).
- AQ5: Please supply captions for figures 1 to 14 (which includes the appendix figures).
- AQ6: Author: Please confirm if this Appendix citation is for Appendix A or Appendix B.
- AQ7: Author: Please confirm if this Appendix citation is for Appendix A or Appendix B.
- AQ8: Author: Please spell out this abbreviation here.
- AQ9: Author: Please spell out this abbreviation.
- AQ10: Author: Please note that we have consecutively numbered the Appendix figures (9–14) as per style. Please check and confirm.
- AQ11: Author: Please spell out this abbreviation at the first instance.
- AQ12: Please provide the place of publication for reference “Arrow, 1962a”.
- AQ13: Please provide university name for references “Beaudry and Francois 2007; Lall 1999; Rhee and Belot 1990”.
- AQ14: The reference “Mankiw et al. 1992” has not been cited in text. Please clarify as to where it should be cited.
-