Managerial Skills Acquisition and the Theory of Economic Development[*]

Paul Beaudry
Department of Economics, UBC, and NBER

Patrick Francois
Department of Economics, UBC, CentER Tilburg, and CEPR

July 2006

Abstract: The aim of this paper is to help understand why not all countries converge rapidly to the use of the most efficient or best practice technologies. Since micro level studies in developing countries suggest managerial skills play a key role in the adoption of modern technologies, and the human resources literature suggests that managerial skills are difficult to codify and learn formally, in this paper we model the interactive process between on-the-job managerial skill acquisition and the adoption of modern technology. The environment considered is one where managerial skills needed to implement best practice technology are learned by observing other competent managers. The paper highlights characteristics that would cause some countries to develop managerial skill and adopt the best practice technologies, while others stay backwards. It also explains why managers may not want to migrate from rich countries to poor countries as would be needed to generate convergence. Finally, we also show why, even with competitive markets, managerial skills may appear to play a small role in cross-country income differences even though they are vital to successful technological adoption.

1

1 Introduction

The process by which technologies that have been successfully implemented in one country are transferred to others, and the reasons why this may not be easy to accomplish, are not yet well understood. In its purest stylized form, familiar from endogenous growth models, technology as a set of instructions for bundling input combinations to produce output should be easy to transmit. However, such an abstracted 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. Learning requires effort, from both teachers and students, resources, and an environment where rewards accrue to those undertaking these efforts. 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".

While both of these mechanisms are of undoubted relevance, 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 little attention. The process we emphasize is a type of "learning by seeing". Like "learning by doing", the mechanism of knowledge transfer we emphasize takes place on the job. With learning by doing, individuals acquire skills and raise the productivity of a new technology by practicing production with it. In contrast, "learning by seeing" is a process whereby, when working as an unskilled worker in production, some of the knowledge that is currently only held by a worker's skilled counterparts is transmitted to the worker through experience. By working alongside (or as subordinate to) the skilled, unskilled workers can succeed in picking up the skills required to fill skilled positions in the firm, or use that experience to themselves start up firms of their own that will replicae the production methods they have learnt. This process of learning both raises productivity directly, and contributes to the future learning of other unskilled workers as they will also be able to work alongside the newly skilled individual and acquire skills themselves. The reason we claim this process is arguably at least as important as the already well studied processes, is because case studies of technology transfer, some of which we elaborate on below, are strongly suggestive of a process like this.

As reported in Easterly (2002), in 1979 Daewoo signed a collaborative agreement with Desh Garment Ltd in Bangladesh. Daewoo agreed to bring 130 Bangladeshi workers to South Korea

1

for training at a Daewoo plant in return for Desh paying commissions amounting to 8 percent of future sales. At the time of the training there were a total of 40 workers producing garments in Bangladesh, and Desh's first year of operations produced $55,050 in sales on 43,000 shirts. By 1987 the industry's output had grown to 2.3 million shirts. During the 1980s, of the 130 Bangladeshi workers initially trained by Daewoo, 115 of them had left Desh to set up their own garment export firms.

By working in the collaborative agreement with Desh, these workers had learned the managerial skills that were the necessary accompaniment to successfully implementing the superior South Korean technology. These workers transported the skills to their own companies and used them to implement the superior technology whose management they had learned on the job. In the process, they too were part of the diffusion of the managerial know-how to their own workers, which in turn further facilitated spread of the technology. By 1985 there were over seven hundred Bangladeshi garment companies. While not knowing exactly how many start-ups were in turn spun-off from those founded by Desh workers, one must assume that it was large – by 1995 Bangladesh was exporting $2 billion in garments (comprising 54% of exports).[1]

This process of on the job skills acquisition, especially managerial skill acquisition, may play a central role in understanding both development successes and processes of technological diffusion. For example, Lall (1999) describes the dramatic increases in the use of managers in the fast-growing East-Asian economies. She also describes the critical role managers played in changing traditional human resource attitudes and policies towards the newer demands required in advanced economies.

Rhee and Belot (1990), in addition to providing an early analysis of Bangladesh's garments industry explosion, have documented the importance of managerial and technical skill transfer in a number of successful case studies drawn from the developing world. An interesting case is the meteoric rise in Indonesian plywood exports through the early 80's that started with two initial firms – Korindo and Kodeco. The take-off points were joint ventures – in the case of Korindo with a Korean lumber exporter, and for Kodeco with one from Taiwan. Part of the knowledge transfer involved the direct relocation of managers from Korea and Taiwan, as real wage increases in both countries had reduced their advantages in labor intensive plywood production. The on-the-job

---

[1]The first detailed case study of Desh was undertaken by Rhee and Belot (1990). Examples of technology diffusion via skilled worker and managerial mobility are commonplace in modern economies. Two examples from advanced sectors are Saxenian (1994), who documents diffusion in Silicon Valley's IT clusters, and Feldman's (2001) study of firm formation in biotechnology and communications industries in Capitol region clusters.

training diffusion process in turn lead to managerial transfer beyond the initial firms. In less than a decade Indonesia moved from being an importer of plywood lumber to accounting for almost 70% of the world's export share – $2 Billion in 1976.

A similar process occurred in the case of Colombian flower exporting. A single company, Floramerica, undertook a feasibility study on exporting flowers to the US in 1969. Prior to that time there were no commercial flower enterprises in the country. Within a year it had 150 employees, and high rates of growth continued. The $100,000 initial investment grew into a company with $50 million in annual sales by 1986. In addition to Floramerica's dramatic growth, the flower producing and marketing know-how diffused to others in the Colombian economy. In less than two decades, flowers became Colombia's fifth largest export, and in the early 90's Colombia was second only to the Netherlands in supplying cut-flowers to the world market. At the period of study, flower production for export directly employed 70,000 people in Colombia, Rhee and Belot (1990, p31).[2]

The means by which the relevant skills are acquired in these examples – on the job – is consistent with evidence from the human resources literature. The know-how that is most important for managing seems difficult to codify and learn through formal institutions. Most managerial learning takes place informally in the workplace – see Enos, Kehrhahn, and Bell (2003) for a recent study and survey of the managerial studies evidence. For example, McCall, Lombardo, and Morrison (1988) found that 30 of 35 managerial job skills were learned informally, with relatively little gained from formal training procedures. The development of proficiency in a skill is a process of tacit learning which seems to depend critically on knowledge gained through experience, and from observing actual workplace leaders at work.[3]

In light of these examples, several questions emerge. For example, how best should one model a process which involves on-the-job learning by seeing as a precursor to entrepreneurial activity and new firm start-ups? How does such a process of technological diffusion (where individuals having experienced the new technology become entrepreneurs or firm managers) compare to alternate processes of diffusion which do not emphasize an initial step of on-the-job learning by seeing? How can the fragility of these examples – whereby a very small number of initially trained individuals appears to have jump started a diffusion process – be explained? Is it necessarily an

---

[2] A more recent study by Piore (2004) estimates that upto 30% of India's IT workforce rotates through American firms to upgrade their tacit knowledge.

[3] Kerr (2005) 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.

3

example of market imperfection? To what extent may such a process of skill acquisition and technology diffusion be an important component on the path to economic development? What role is played by migration, and what are the private incentives for skilled individuals to migrate to countries where their skills are rare? Can such a process that emphasizes the benefits of experience be reconciled with the relatively small observed estimates of returns to experience found in empirical studies?

In this paper we present a model of the interactive process between managerial skill acquisition and the adoption of new technology that tackles these questions. We use the model to help isolate factors that may explain why certain countries may adopt a new technology, while others would stay backwards. There are two important aspects to our modelling strategy. First, we model the acquisition of (technology specific) managerial skill as resulting from on-the-job interaction with knowledgeable managers. In particular, as an individual works in a firm and interacts with competent managers, he can gain the skills of a manager and potentially open his own firm. However, in contrast to much of the literature on learning-by-doing, the skills obtained as a manager are not assumed to be a substitute to his previous skills, but instead are modelled as a complement to such previous skills. That is, managerial skill is not about getting better at what one previously did, but instead is modelled as learning how to organize non-managerial workers and thereby increase their productivity. As we shall show, this difference in the modelling of the learning-by-doing process has strong implications for the dynamics of managerial skill premia. Second, we assume that all the learning activities are internalized by the market in the sense that that workers can accept lower wages to be hired in firms where they are likely to obtain managerial skills (as in Becker (1975)). Hence, in our set up, there will be no labor market imperfections. This assumption is not chosen for its realism, but instead to isolate implications of the process of managerial skill acquisition that are independent of implications of market imperfections. As a result of this assumption, the market will generate a hedonic type equilibrium, where jobs bundle both wage payments and managerial learning possibilities.[4]

Our analysis of managerial skill acquisition gives rise to two main insights. On the one hand, we show that this process of acquisition shares many similarities with the process of technological innovation emphasized in the endogenous growth literature, even though our environment is characterized by the absence of any market imperfections or non-convexities, and that different

---

[4]In ongoing work, we are exploring how the results here are modified and amplified in the presence of market imperfections such as in the presence of credit constraints, information asymmetries or thin market effects.

technologies are freely available. For example, in our model, agents will purposely accumulate managerial skill knowing full well that the returns to these skills will dissipate over time, and possibly go to zero. The reason why the price of managerial skill drops as the new technology is adopted is because knowledgeable managers are becoming more abundant. Moreover, the managerial wage premia may go to zero due to the fact that the learning process may continue even if skills are abundant. Hence, in such a case, the incentive to invest in managerial skill will be entirely driven by expected temporary rewards, which is akin to the incentive to innovate in a R&D model with finite lived patents.

The second interesting finding of the analysis is that the incentive for managers to migrate from a more developed country (where managerial knowledge is abundant) to a less developed country (where it is scarce) is shown to be very weak and possibly non-existent. This observation has the potential to help understand the persistence of under-development since, contrary to most neo-classical models, it does not predict that the scarce factor will want to migrate from richer to poorer countries. The reason being that the wages of managers can be higher in the country where managers are more abundant, due to their teaching value, even though technological opportunities satisfy constant returns to scale.

At a first pass, it may appear that the literature on the sources of cross-country income differences (for example, Klenow and Rodriguez-Clare 1997) rules out the type of managerial knowledge studied here as an important element in development. In particular, this literature finds that differences across countries in the amount of human capital obtained on-the-job are a very minor element in explaining cross-country differences in income. In this literature, the value of skills obtained on the job are proxied by the returns to experience, and these do not seem to vary sufficiently across countries to be able to explain a large part of cross country variation in income. Since managerial skills should be a subset of on-the-job (experience driven) human capital, this a fortiori appears to rule out managerial skills (and the institutions that favor managerial skill) as an important element driving cross-country differences in income. However, as we shall show, this reasoning is flawed when a process of learning by seeing underlies the transmission process. In particular, we will show why an arbitrarily large difference in income could be due to differences in available managerial skill across countries with access to the same technology. Standard methods used for accounting for cross-country income differences would wrongly attribute all the difference to measured total factor productivity; thereby suggesting that the differences are due to different

availability of technologies.[5]

This paper relates to both the literature on learning-by-doing and the literature on technology diffusion. In particular, it relates to the learning by doing literature since managerial skills are assumed to be acquired on the job.[6] However, as mentioned above, the key differences between our paper and most of the learning-by-doing literature are that we do not model learning as inducing higher productivity in the same task, but instead model the learning process as allowing one to obtain skills that are used to organize non-managerial workers efficiently in a new technology. [7].

The paper proceeds as follows. Section 2 sets up a model examining managerial skills acquisition in a competitive economy. The steady state and transitional dynamics of such a process are analyzed, with an emphasis on how the equilibrium properties are affected by structural (preference) parameters. The resulting mapping between parameters and outcomes offers an explanation for why some countries may adopt a technology while others don't. It also explains why a small perturbation to initial condition may set off the process of technological adoption. Section 3 explores the model's implications for migration. In particular, this section shows why unskilled labor that moves from poor to rich countries may reap high returns, whereas similar incentives may not exist for the movement of skilled labor to poor countries. Section 4 examines the consequences of this process for growth accounting. It is shown that these fully internalized on-the-job skills need not be detectable in cross-country growth accounting exercises, and if showing up at all, are extremely unlikely to reflect their true value. Thus our model is completely consistent with observed returns to experience varying little across countries. Section 5 concludes.

---

[5]Growth theories that emphasize non-rivalrous technology (such as Arrow (1962a), Shell (1966), Grossman and Helpman (1991) and Aghion and Howitt(1992)) need to explain why such technologies do not diffuse rapidly to poor countries. There are many proposed explanation to this puzzle. For example, one such explanation is the presence of "barriers to adoption", such as legal, institutional, socio-cultural or political factors, see Parente and Prescott (2002). Another class of explanations emphasizes difficulties in absorption of new technologies. These could be due to differences in skill levels, Acemoglu and Zilibotti (2001) which may interact with within firm contracting imperfections – Aghion, Acemoglu and Zilibotti (2004). This paper offers an alternative explanation based on the process of accumulation and diffusion of the managerial skills needed to use a new technology.

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

[7]In relation to the literature on technological diffusion, the paper closest in spirit is that of Chari and Hopenhayn (1991). Their study analyzed rates of technology diffusion in a model with continuous technology upgrading and vintage specificity of skills in production. In a 2-period lived overlapping generations economy where the set of possible technologies expands each period, they demonstrate the existence of a stationary distribution of technologies in use (with constantly increasing average productivity) and constantly increasing wages. Their analysis 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.

6

2 The Model

Households

Consider an economy where the population (or labor force) is of constant size and normalized to 1. At each instant a new cohort of size δ enters and a measure δ of workers dies; with the property that $\int_{-\infty}^{t} \delta e^{-\delta(t-s)} ds = 1$. Each individual is risk neutral and inelastically supplies one unit of labor per instant. Let $\rho > 0$ be the instantaneous discount rate. All newly entering workers do not have managerial skills, and are thus unskilled. Discounted expected lifetime utility for an individual is:

$$U = \int_{0}^{\infty} c(t) e^{-(\delta+\rho)t} dt,$$

where $c(t)$ denotes consumption.[8]

Production

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

Total output produced using S skilled managers, U unskilled workers, and L laborers is given by:

$$Y = F(S,U) + H(L)$$

and since H is linearly homogeneous:

$$Y = F(S,U) + \overline{w}L,$$

where $\overline{w}$ is labor's marginal product in the traditional technology. [9] We impose the following

---

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

[9] Note that capital plays no role in the analysis and that this is without loss of generality.

Inada type condition on production:

$$
\begin{aligned}
\text{for } U &> 0, \quad \lim_{S \to 0} F_1(S,U) = \infty \text{ and } F_2(0,U) = 0 \\
\text{for } S &> 0, \quad \lim_{U \to 0} F_2(S,U) = \infty \text{ and } F_1(S,0) = 0.
\end{aligned} \tag{1}
$$

$$
\text{by continuity then } \exists \, S' \quad : \quad F_1\left(S', 1 - S'\right) = F_2\left(S', 1 - S'\right). \tag{2}
$$

A natural interpretation of this last condition is that production necessitates some individuals being placed in positions requiring organizational skills, i.e., management, but when all such positions are occupied, extra workers with such skills are more usefully allocated to positions that could as easily be filled by the unskilled. When this occurs, the productivity differential between a skilled manager and an unskilled worker disappears.[10]

Given that we want the modern technology to possibly take over from traditional technology we assume that, if the modern technology is not managerial skill constrained, the marginal unskilled individual is more productive working with it than in the traditional technology, i.e.:

$$
F_1\left(S', U'\right) = F_2\left(S', U'\right) > \overline{w}. \tag{3}
$$

On-the-job acquisition of managerial skills

If an unskilled individual works in the modern technology, we assume that he acquires managerial skills randomly following a Poisson process, where the instantaneous rate of learning, denoted $\Omega(S,U)$,, depends on both the number of managers and the number of workers in a firm.[11] In particular, we assume that $\Omega_1 \geq 0$, $\Omega_2 \leq 0$ and Ω is homogeneous of degree zero. The homogeneity of degree zero implies that the speed of learning in a firm depends only on the ratio of managers to workers. We will also employ the notation $\widetilde{\Omega}(S,U)$ to denote the function $\Omega(S,U)U$ which is the instantaneous output of the skill production process. Clearly $\widetilde{\Omega}(S,U)$ is homogenous of degree one and we further impose the restriction that it is a concave function.

In order to allow the process of adoption of the modern technology to begin with $S = 0$, we assume that managerial skills can be learned by working with the new technology even without

---

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

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

the presence of other managers, that is, we assume that $\Omega(0, x) > 0$ for $x > 0$. A skilled individual remains skilled throughout their life - i.e., individual skills do not depreciate.

2.1 The Decentralized Equilibrium

In the above setup, managerial skills acquisition is a by-product of production. However, this does not imply that there are externalities in this process since skills acquisition is excludable. Individuals who do not work under skilled managers in the modern technology cannot acquire managerial skills. This teaching aspect of managers allows firms with a high ratio of managers to offer a low wage to unskilled workers knowing that the offer may still be attractive since an unskilled worker can learn managerial skills more quickly in such a firm. In such an environment, profit maximizing firms make employment decisions taking into account that their ratio of managers to unskilled workers affects the unskilled wages they can offer. Hence, an employment offer to an unskilled individual is actually a bundle that specifies a wage, w_t^u and a (firm specific) speed of learning Ω. Among bundles of w^u and Ω, an unskilled worker will accordingly choose the one that maximizes his utility. Our objective will be to characterize the situation where both firms and workers take the market tradeoff between w^u and Ω as given when making their decisions, with the equilibrium requirement that demand equals supply at each chosen pair of w^u and Ω. This type of competitive equilibrium is known as a hedonic competitive equilibrium (see Rosen (1974)). In our case, we want to analyze a dynamic hedonic equilibrium where there is a sequence of tradeoffs between w^u and Ω, one at each moment in time. For further reference, it will be useful to denote the market tradeoff between w^u and Ω at time t by the function $g_t(\cdot)$, with

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

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

2.1.1 The Firm's Problem in a Hedonic Equilibrium

Firms are competitive, and can costlessly enter and exit instantaneously. A firm's objective is thus the simple static problem of maximizing profits by choosing the number of skilled workers to place in managerial positions, the number of unskilled workers, and the wage to pay unskilled workers

w_t^u taking as given the market wage of skilled workers, w_t^s, and the market tradeoff $g_t(\cdot)$ between w^u and Ω. A firm hiring S_t skilled workers places a fraction $\gamma_t \leq 1$ into managerial positions, and the remainder $(1 - \gamma_t) S_t$ into unskilled positions where they work with the unskilled workers, U_t. This maximization is therefore given by the following, where firms recognize how their choice of S_t and U_t affects the speed of learning $\Omega(\cdot)$:

$$\max_{\gamma,S,U,w^u} F(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - w_t^s S_t - w_t^u U_t$$

$$s.t.$$

$$w_t^u = g_t(\Omega(S_t, U_t)), \text{ and } \gamma_t \leq 1.$$

This yields first order conditions:

$$w_t^s = F_1(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - g_t'(\Omega) \Omega_1(S_t, U_t) U_t \tag{4}$$

$$w_t^u = F_2(\gamma_t S_t, (1 - \gamma_t) S_t + U_t) - g_t'(\Omega) \Omega_2(S_t, U_t) U_t \tag{5}$$

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

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 the Hedonic Equilibrium

Since consumers are risk neutral, it is easiest to analyze the household's labor allocation problem at the level of the cohort rather than at the level of the individual. In effect, instead of first determining optimal individual decisions within each cohort and then aggregating them up, the assumption of risk neutrality allows us to first aggregate the individual into the cohort and then derive the optimal decisions for the cohort. This equivalence is rather well known, so we omit the proof. The relevant cohort is still small, so they act as price takers in their allocation decisions and the problem is best stated as one where members pool income so that each member has equal consumption.[12] The problem then amounts to allocating skilled and unskilled workers between production in both the traditional and modern technology at each instant in order to maximize per capita consumption. Cohort q takes as given its stock of individuals with managerial skill at

---

[12] Since individuals are risk-neutral, the solution for the pooled cohort is equivalent to that for each individual.

time t, denoted $S(q,t)$. These individuals receive the market wage for skilled, w_t^s. Though it is possible for these individuals to remain in the old technology receiving $\overline{w}$, since this is never part of an optimal program for the skilled, to save on notation, we do not denote this possibility.

The cohort's unskilled workers can work in the new technology (where they probabilistically acquire managerial skills) and receive wage w_t^u. This proportion is given by $\beta(q,t)$. Alternatively, unskilled workers can work in the old technology receiving $\overline{w}$ and remain unskilled for sure, with this proportion denoted $(1 - \beta(q,t))$.

Denote the cohort's remaining size at t as $N(q,t)$ so that its measure of unskilled is $N(q,t) - S(q,t)$. The cohort's optimization problem is:

$$\max_{\beta,S,\Omega,w^u} \int_q^\infty \left(\frac{C(q,t)}{N(q,t)}\right) e^{-(\delta+\rho)(\tau-q)} d\tau$$
$$s.t.$$

$$C(q,t) = w_t^s S(q,t) + w_t^u \left[(N(q,t) - S(q,t))\beta(q,t)\right]$$
$$+\overline{w}\left[(1 - \beta(q,t))(N(q,t) - S(q,t))\right] \tag{7}$$

$$w_t^u = g_t(\Omega) \tag{8}$$

$$\dot{S}(q,t) = (N(q,t) - S(q,t))\beta(q,t)\Omega - \delta S(q,t) \tag{9}$$

$$N(q,t) = e^{-\delta(t-q)} \tag{10}$$

$$\beta(q,t) \in [0,1] \; \forall t.^{13} \tag{11}$$

A simple transformation of variables proves useful in analyzing this problem. Let the fraction of the cohort that are skilled be denoted $\chi(q,t) = \frac{S(q,t)}{N(q,t)}$ and $c(q,t) = \frac{C(q,t)}{N(q,t)}$, then:

$$\frac{\dot{\chi}(q,t)}{\chi(q,t)} = \left(\frac{1}{\chi(q,t)} - 1\right)\beta(q,t)\Omega \tag{12}$$
$$\Rightarrow \dot{\chi}(q,t) = (1 - \chi(q,t))\beta(q,t)\Omega \tag{13}$$

$$and$$

$$c(q,t) = w_t^s \chi(q,t) + w_t^u (1 - \chi(q,t))\beta(q,t)$$
$$+\overline{w}(1 - \beta(q,t))(1 - \chi(q,t)). \tag{14}$$

The cohort's problem can therefore be rewritten as:

$$\max_{\beta,\chi,w^u,\Omega} \int_0^\infty c(q,t) e^{-(\delta+\rho)(\tau-q)} d\tau$$
$$s.t. \; (8) \; (11) \; (13) \text{ and } (14).$$

11

The first order conditions of this problem imply:

$$\overline{w} - w_t^u \begin{cases} > \lambda_t \Omega & \text{if } \beta_t = 0 \\ = \lambda_t \Omega & \text{if } \beta_t \in (0,1) \\ < \lambda_t \Omega & \text{if } \beta_t = 1 \end{cases} \tag{15}$$

$$g_t'(\Omega) = -\lambda_t \tag{16}$$

$$\dot{\lambda}_t - (\delta + \rho)\lambda_t = \beta_t w_t^u + (1 - \beta_t)w_s^o - w_t^s + \lambda_t \beta_t \Omega$$

and the transversality condition is:

$$\lim_{t \to \infty} e^{-\rho t} \lambda_t S_t = 0. \tag{17}$$

Intuitively, the cohorts main tradeoff is the tension it faces between allocating its unskilled workers to working in the new technology at a low wage, w_t^u, but in return probabilistically learning skills that can have a positive shadow value, λ_t, that will allow a future stream of higher wages w_t^s. The function $g_t(\Omega)$ describes the tradeoff imposed on the cohort by the hedonic equilibrium. A higher probability of learning these skills, i.e., a firm with a high ratio of managers to workers, pays a correspondingly lower wage, with the amount of discount positively related to the shadow value of skills, λ_t as described by (16).

Integrating across cohorts yields:

$$S_t = \int_0^t \chi(q,t) N(q,t) \, dq$$

$$\Rightarrow S_t = \int_0^t \chi(q,t) e^{-\delta(t-q)} dq$$

$$and$$

$$\dot{S}_t = (1 - S_t) \beta_t \Omega - \delta S_t. \tag{18}$$

Substituting for wages from (4) and (5), the problem of characterizing the market equilibrium

12

can be reduced to characterizing the following system in β_t, γ_t, S_t and λ_t.

$$\text{Either } F_1(S_t, 1 - S_t) - F_2(S_t, 1 - S_t) \geq 0 \text{ and}$$

$$\gamma = 1$$

$$\overline{w} - F_2(S_t, \beta_t(1 - S_t)) \begin{cases} > \lambda_t \Omega & \text{if } \beta_t = 0 \\ = \lambda_t \Omega & \text{if } \beta_t \in (0, 1) \\ < \lambda_t \Omega & \text{if } \beta_t = 1 \end{cases}$$

$$\dot{\lambda}_t - (\delta + \rho)\lambda_t = \beta_t F_2(S_t, \beta_t(1 - S_t)) + (1 - \beta_t) w_s^o - F_1(S_t, \beta_t(1 - S_t))$$

$$+ \lambda_t \beta_t \Omega(S_t, \beta_t(1 - S_t)) \tag{19}$$

$$\text{OR}$$

$$\gamma < 1, \lambda_t = 0, \beta_t = 1 \text{ and}$$

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

with the addition that the transversality condition (17), and the aggregate accumulation equation (18), must hold.[14]

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

2.2 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 (3)). 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]

---

[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.

[15]Note that potential managers never choose the traditional technology since the modern technology is superior.
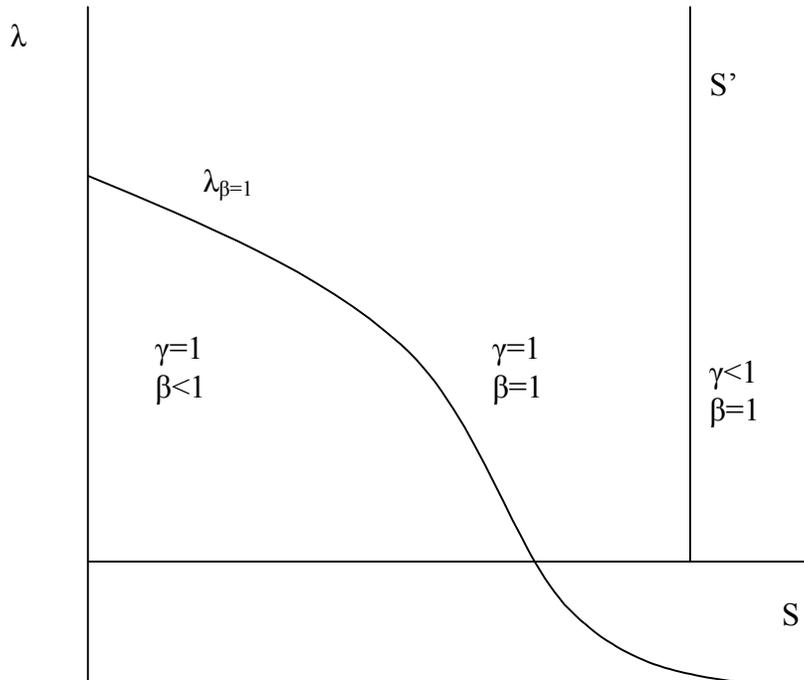
Figure 1

The second line in Figure 1 depicts the set of points that delimit the region where $\beta = 1$ versus when $\beta < 1$. This relationship can be expressed as $\lambda = \frac{\overline{w} - F_2(S,(1-S))}{\Omega + \Omega_2(1-S)}$ and is derived from condition (15). We will denote this relationship as $\lambda_{\beta=1}$ from now on. In the appendix it is proved that this schedule is downward sloping. 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 represents a region where it is optimal to leave some unskilled workers working with the traditional technology.

The $\lambda_{\beta=1}$ line plays a particularly important role since the properties of the $\dot{S} = 0$ line and the $\dot{\lambda} = 0$ line differ depending on whether they lie in the $\beta < 1$ region or the $\beta = 1$ region. The next step is therefore to characterize the dynamics of S and λ depending on whether the $\dot{S} = 0$ line or the $\dot{\lambda} = 0$ line cross the $\lambda_{\beta=1}$ locus. The details of this characterization are found in the appendix with the salient results regarding the shape of the $\dot{\lambda} = 0$ and $\dot{S} = 0$ lines summarized below:

Lemma 1: *Consider the* $S - \lambda$ *space, as depicted in Figure 1.*
For $S - \lambda$ *values such that* $\beta = 1$ *:*
(1) The $\dot{\lambda} = 0$ *locus is negatively sloped.*

14

(2) The $\dot{S} = 0$ locus is vertical.

    For $S - \lambda$ values such that $\beta < 1$:

(3) The $\dot{\lambda} = 0$ locus is horizontal.

(4) The $\dot{S} = 0$ locus is horizontal.

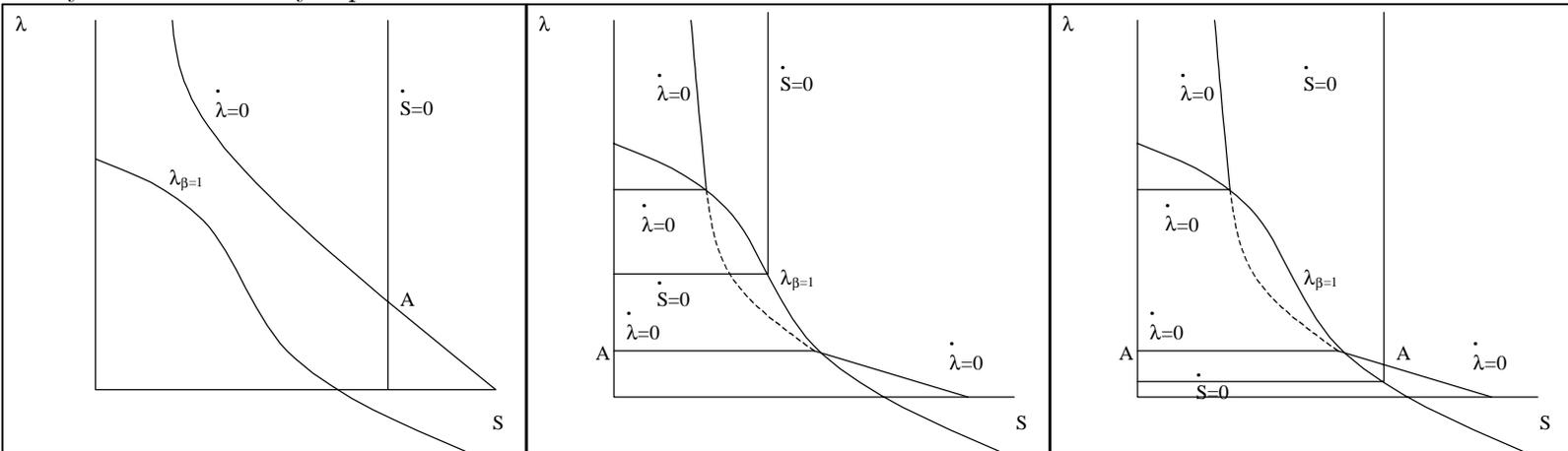## 2.3   Steady States and Transitional Dynamics

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 occuring with the modern technology and one of which has all production occuring with the traditional technology.

Somewhat surprisingly a hybrid steady state where both modern and traditional production methods can 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 2 below, with steady states denoted by capital A.

A. Unique Modern Steady State   B. Unique Traditional Steady State   C. Modern and Traditional Steady State

Figure 2

To better understand the implications of these configurations we need to look at the associated transitional dynamics. Figure 3 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 6, the unique steady state is characterized by a positive managerial skill premium (since the steady state value of λ is positive). This is due to the fact that the vertical $\dot{S}$ 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, 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.
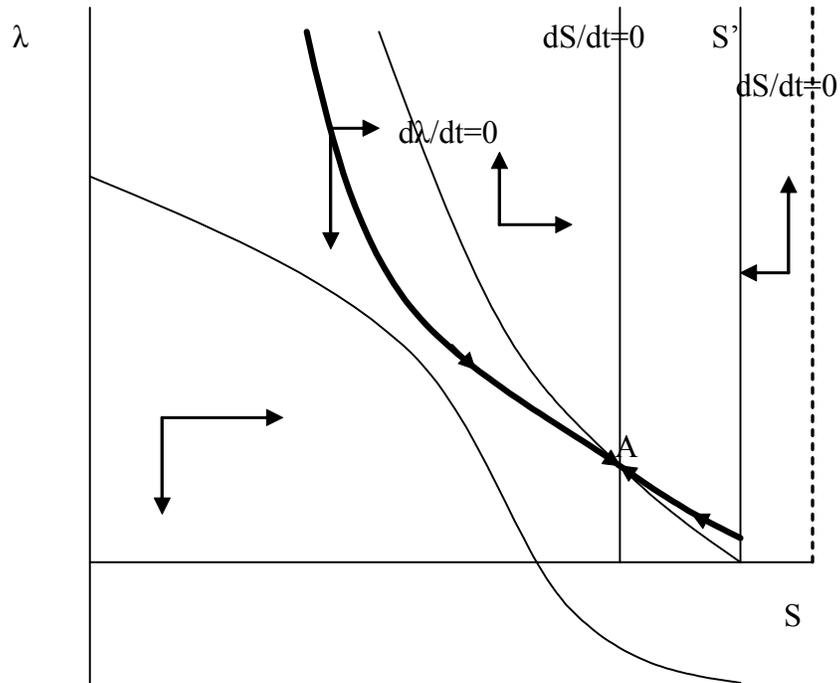

λ   dS/dt=0   S'
                                       dS/dt=0

                    dλ/dt=0

                                    A

                                          S

Figure 3

16

Figure 4 depicts the phase space for the situation corresponding to crossing of both loci with the $\lambda_{\beta=1}$ line. Here again, there is a unique steady state and it is 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 cohort turnover rate. In such an economy, skill premia in the modern technology will be steadily rising as it approaches the 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, λ increases as the transition path is heading in a northeasterly direction, and all individuals are being placed into learning the new technology, the economy's skill level is falling. Eventually, the transition path enters the $\beta < 1$ region, since the lower numbers of skilled managers would lower the wages of the unskilled below opportunity cost if all were placed into learning the new technology. When this occurs, the erosion of the skill stock is even greater and the economy converges alont the $\dot{\lambda} = 0$ locus towards the steady state at A.

λ

dλ/dt=0   dS/dt=0   S'
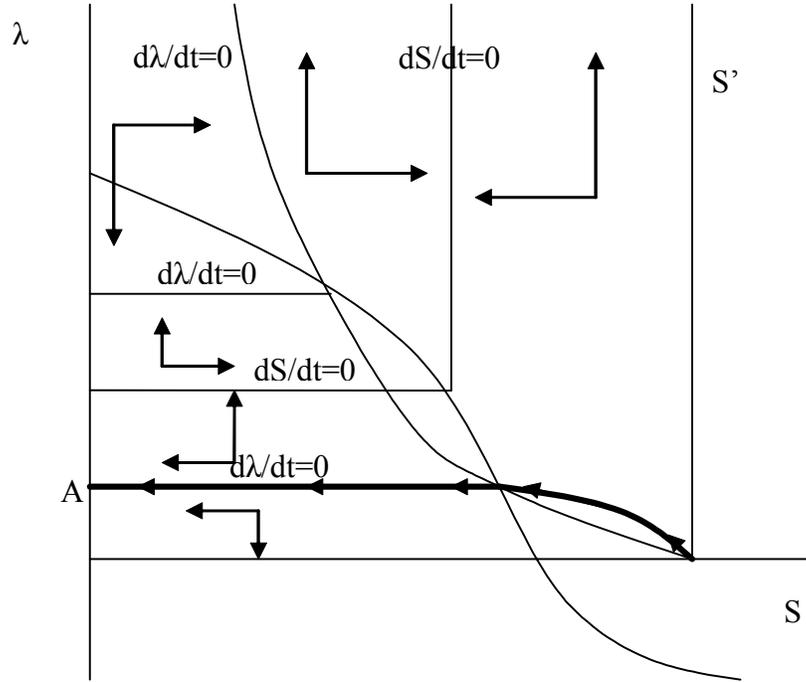
dλ/dt=0

dS/dt=0

A   dλ/dt=0

S

Figure 4

Figure 5 depicts the situation that occurs when only the $\dot{\lambda} = 0$ locus intersects 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 labelled 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. But then still more unskilled workers will 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

18

$\dot{S} = 0$ line denoted $\dot{S} = 0(2)$. This occurs when the $\dot{S} = 0$ locus does not intersect the $\lambda_{\beta=1}$ line.[16]
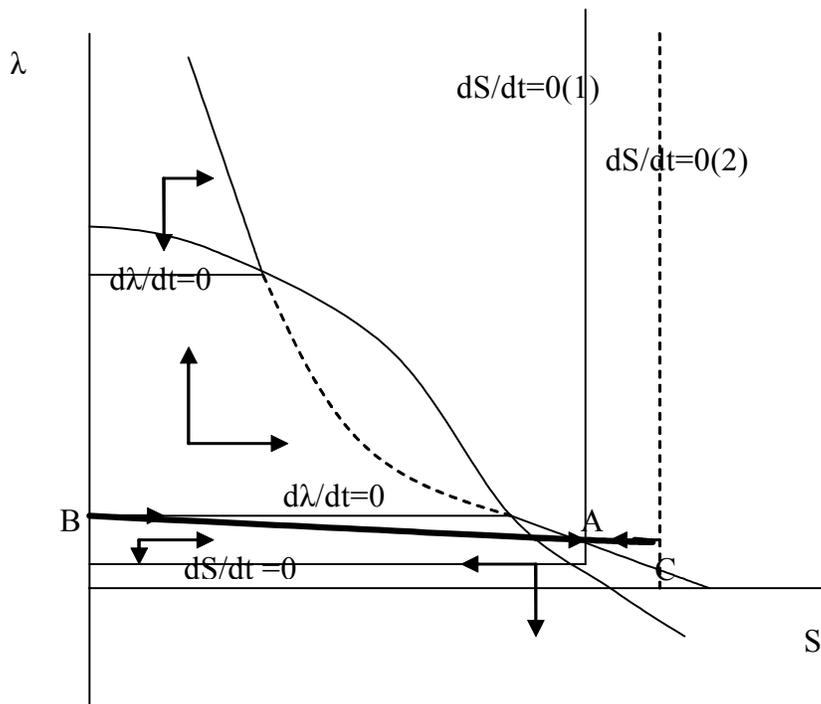

λ

dS/dt=0(1)

dS/dt=0(2)

dλ/dt=0

dλ/dt=0

B

A

dS/dt =0

C

S

Figure 5

The transition paths depicted above highlight a form of monotonicity in convergence patterns as stated in Proposition 2.

Proposition 2: *Technology adoption is monotonic. For any initial level of managerial skill, S, if, at any time, the proportion of producers using modern technology falls, then all modern production eventually ceases. If, at any time, the proportion of producers using modern technology rises, then all production eventually uses modern technology.*

It is not possible for an economy to intensively invest in accumulation of managerial skills early in the introduction of a new technology and then reverse itself as the technology becomes more widespread. This is because the transition path to any steady state cannot enter and re-emerge from the $\beta < 1$ phase. If it is ever worthwhile to leave some workers in the old technology, then all workers will eventually work there. If it is worthwhile to place some workers into learning the

---

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

new technology, when there are few skills present, then it will also be worthwhile to have them work in the new technology when skills are widespread.

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

Proposition 3: *For a given productivity level in the modern technology, F, and a given managerial learning technology Ω the $\rho - \delta$ space can be divided into three regions*

(a) For low values of ρ and δ, denoted region A, the economy converges to a unique steady state with all production occurring in the modern technology.[17]

b) For high values of ρ and δ, denoted region B, the economy converges to a unique steady state with all production occurring in the traditional technology.[18]

c) For intermediate values of ρ and δ, denoted region C, if the economy starts with $S = 0$, it remains there with all production occurring in the traditional technology. If however the economy starts with $S > 0$, it converges to a steady state with all production occurring in the modern technology.

Furthermore, there exists a δ^*, such that if $\delta < \delta^*$, then there is no managerial skill premium in a steady state where all production occurs with the modern technology.[19]

Figure 6 summarizes this proposition's implications for possible steady state outcomes. Note that Region A corresponds to the configuration in Figure 3, region B corresponds to the configuration in Figure 4 and region C corresponds to the configuration in Figure 5.

In the lower left corner of Figure 6 (i.e. low values of δ and ρ) lies the region where an economy with an arbitrary level of initial skills (including zero) will fully transform to modern production. If in this case $\delta < \delta^*$, there will be no managerial premium in the steady state. The disappearance of a managerial premium follows directly from the low labor turnover that ensures steady state skills are high. For higher values of δ, while staying in Region A, the modern technology will always fully develop, but the higher values of δ will allow the persistence of a skill

---

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

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

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

premium in the steady state. For high values of both δ and ρ, as in Region B, development of the modern technology is not sustainable. An economy inheriting some skilled managers, or receiving in-migration of them, would utilize the technology, and would also train some further managers in its use. However, the low rate of skill creation is not sufficient to offset the high labor turnover and therefore such an economy will eventually 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.
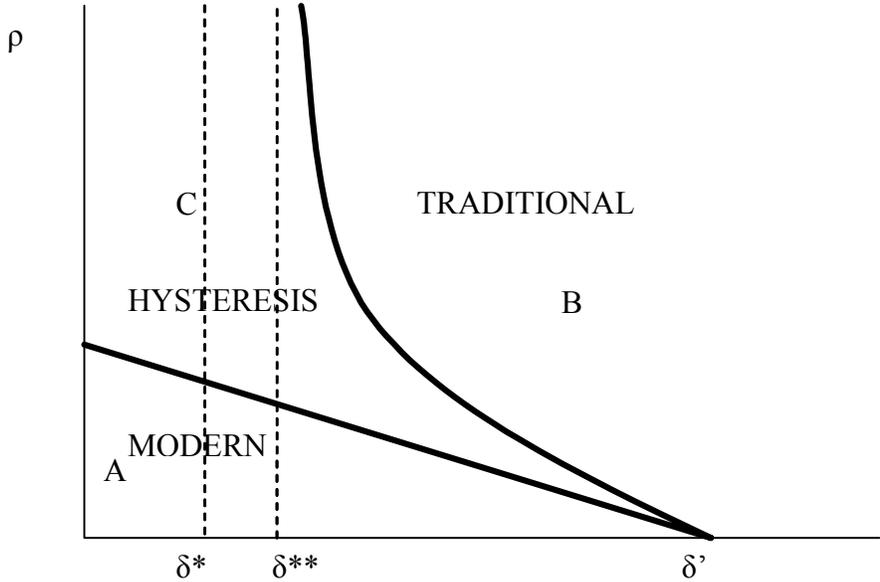

Figure 6

Though on the job skills will be accumulated in all economies, the model provides an interpretation of when these will lead to a take-off into the new technology, and thus a growth miracle, and when efforts at implementing new technologies will fail. Specifically, as illustrated in Figures 3, even an economy without any managerial skills will devote resources to acquiring the new

technology since both δ and ρ are sufficiently low. Starting at $S = 0$, growth rates will be high, and returns to experience will be large at the start and eventually decline, perhaps to zero, in the eventual steady state. The factor in common in both the figures is a high valuation of the future. In extreme contrast, Figure 4 depicts an economy with both a relatively low valuation of the future, and high turnover. Such an economy would also experience some growth were it to receive an infusion of skills from abroad. However, even a large injection of skilled workers would not transform this economy to modern production, instead, it will converge back to the traditional technology, along a path which sees constantly declining output. The middling case is given in Figure 5. Such an economy will not converge to modern production without some exogenous influx of skills from outside the system. Any initiative which does this, though, will lead to rapid learning, diffusion, and sustained growth to the new steady state where there is a permanently higher level of income.

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 flows. This arises for standard neo-classical reasons.[20] Since skills are relatively abundant in developed countries, returns to them should be low there, and there should thus be incentives for skilled individuals to migrate to less developed countries where their skills are scarce. In the present framework, skills are embodied and fully rewarded, suggesting that a similar set of incentives should be in place. It turns out, however, that this is not generally the case.

Here we shall consider migration incentives between two countries with access to the same set of technologies, modern and traditional, but differing in one of three ways that will generate dramatic outcome differences. The first two sources of difference are fundamentals: the discount factor, ρ; and the population turnover rate, δ. As we have seen previously, countries with low values of each of these are more likely to converge to steady states where the modern technology dominates. Conversely, countries with higher values of these variables are more likely to be in a steady state where all production occurs using the lower productivity traditional technology. Finally, we consider migration incentives implied by the hysteresis case illustrated in Figure 9 where fundamentals are identical.

---

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

When considering incentives to migrate, we focus on instantaneous incentives. If a wage payment is higher – for a given skill level – in country i versus country j, we will say that such a skilled worker has an incentive to migrate from country j to country i. Our primary focus is on the incentive for skilled managers to migrate from richer to poorer countries. We omit discussion of the migration incentives for unskilled workers since they are quite standard in the model in the sense that there is never an incentive for an unskilled worker to migrate from a rich country to a poor country.

3.1 Migration Incentives due to Differences in ρ

Consider two countries, rich, r, and poor, p, with identical values of δ but with differing discount rates such that $\rho^r < \rho^p$. We choose values of the respective ρ so that the rich country is in Region A in Figure 6, while the poor country is in Region B. Note that such a configuration is only possible if $\delta > \delta^*$.

This choice of ρ ensures that the poor country will be in a steady state in which all production occurs in the traditional technology, and that the rich country will be in a steady state where all production is in the modern technology. Also, in the rich country's steady state there is ongoing training of individuals in managerial skills, whereas in the poor country's steady state, managerial skills are not being produced as no one works with the modern technology. This does not mean that managerial skills would not be valuable in the poor country. Specifically, if one skilled manager from the rich country, where skills are widespread, were to migrate to the poor country, he would be able to set up production and utilize the modern technology. Moreover, the unskilled workers he employed would also acquire valuable managerial skills and therefore be willing to "pay" for these by working at a wage below the wage prevailing with the old technology. This would generate a positive return to managerial skill and would seem to suggest strong incentives for migration of skilled individuals from the rich country to the poor

one. It turns out that this is not the case.[21]

Proposition 4: *Income differences generated by differences in ρ across countries do not generate incentives for managers in a richer country to migrate to a poorer country.*

Managers in the rich country do not want to migrate to the poor country because the wage they receive is higher in the rich country (i.e., low ρ country). But how can that be given that managers are more abundant in the rich country? The answer lies in the teaching value of managerial skills. Recall that wage payments to a manager are made up of two components: the marginal product in production and the teaching value. The lower value of ρ, which makes the country richer, does this by inducing more individuals to accumulate managerial skills. This ensures a high teaching value of managers and since many more unskilled workers are willing to learn managerial skills, this guarantees overall managerial productivity, and hence wages, are higher in the rich country. This finding is consistent with the evidence presented in Acemoglu and Newman (2002). They report averages of the ratio of managerial to production workers in six OECD countries, all of which are below 25%. In contrast, the average supervision ratio in Sub-Saharan Africa is 41%. The explanation provided here would be that the lower steady state value of managerial skills (due to the higher rate of discount, ρ) in the Sub-Saharan economies implies that workers are less willing to pay the cost of being a low productivity worker required to obtain them.

3.2 Migration Incentives due to Differences in δ

Now consider the induced migration incentives between two countries that differ in terms of δ, but have the same value of ρ. The first county has a high value of δ, denoted δ^p, placing it in region B of Figure 6 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

---

[21] The managerial wage in the poor country is given by

$$w^s(\delta) = F_1(Z,1) + \lambda \widetilde{\Omega}_1(Z,1), \tag{20}$$

where λ is given by

$$\lambda = \frac{F_1(Z,1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z,1)},$$

and where Z is the smallest root of:

$$\frac{F_1(Z,1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z,1)} = \frac{\overline{w} - F_2(Z,1)}{\widetilde{\Omega}_2(Z,1)}.$$

6.[22] 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. In order to understand these in Figure 7 we plot the wage of skilled workers as a function of δ. Recall that their wages can be expressed as follows:

$$w^s\left(\delta\right) = F_1\left(Z,1\right) + \lambda\widetilde{\Omega}_1\left(Z,1\right), \tag{21}$$

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 4, there is a well defined steady state value for λ[23] 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:

$$\lambda = \frac{\overline{w} - F_2\left(Z,1\right)}{\widetilde{\Omega}_2\left(Z,1\right)}.$$

For the case where the steady state is characterized by exclusive use of the the modern technology, then the relvant values for Z and λ are determined by the $\dot{S} = 0$ and $\dot{\lambda} = 0$ conditions when $\beta = 1$.

---

[22] If a country has a parameter configuration which places it in Region C of Figure 10, we consider here that such a country is in the high output steady state.

[23] This λ is given by

$$\lambda = \frac{F_1\left(Z,1\right) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1\left(Z,1\right)},$$

where Z is the smallest root of:

$$\frac{F_1\left(Z,1\right) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1\left(Z,1\right)} = \frac{\overline{w} - F_2\left(Z,1\right)}{\widetilde{\Omega}_2\left(Z,1\right)}.$$
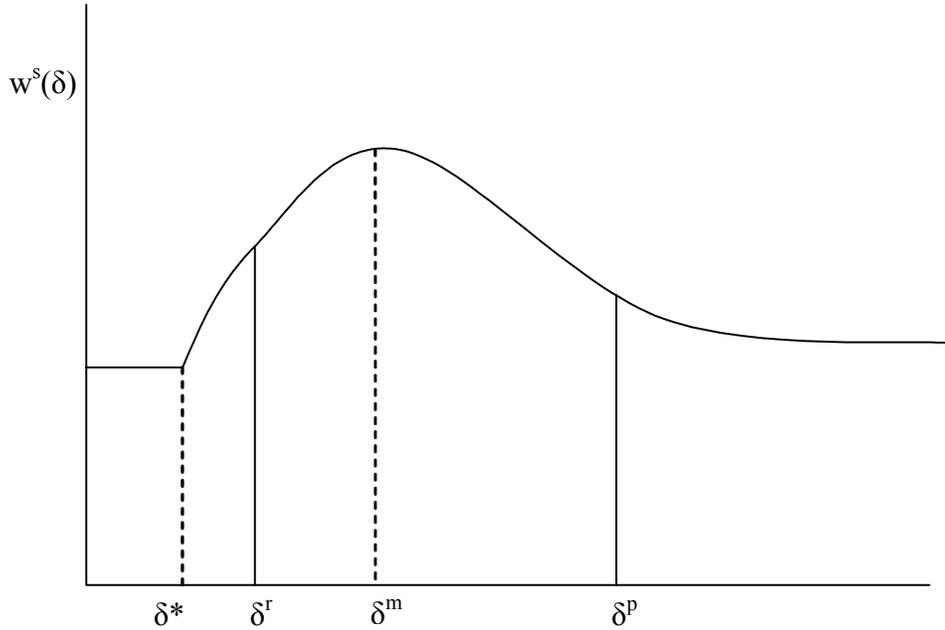
Figure 7

The important aspect to notice in Figure 7 is that 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 6. 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 of the reduced teaching value of being skilled. Higher turnover implies the unskilled are less willing to 'pay' to learn managerial skills through low wages so that skilled wages are falling in δ. 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 forces dominate so that wages for the skilled in steady state will be lower the greater the proportion skilled, i.e., the lower is δ. The non-monotonic relationship between w^s and δ implies that a manager in a developed/rich country does not necessarily want to migrate to a poor country where his skills are less abundant. In fact, for any given poor country, there exist a range of δs (with $\delta < \delta^m$) such that, a country in this range is characterized (in steady state) by both the exclusive use of the modern technology and the absence of any incentive for the skilled to migrate to the poorer country. This observation is summarized in Proposition 5. Note however, that if the δ of the rich country is sufficiently small, then a skilled manager will have

26

an incentive to migrate to the poor country; since for a sufficiently small δ in the rich country, the managerial wage is lower than in the poor country. Once again, the reason why a manager in a rich country may not want to migrate to the poor country is because the teaching value of having skill is greater in the rich country.

Proposition 5. *Cross-country income differences generated by differences in δ generate incentives for skilled managers to migrate to poor countries only if δ in the rich country is sufficiently low.*

It may seem odd to treat as country level primitives the parameters ρ and δ when allowing individuals, who embody such primitives, to relocate. Presumably, if enough individuals from the low ρ or δ country migrate to the high one then there is a force for homogenization of these values. However this is not really a problem for our propositions 4 and 5 above. The reason is that these explain why skilled managers will not find it worthwhile to migrate from countries where skills are abundant to ones where they are scarce, even with potentially large differences in parameters. Consequently, since they explain cases under which there will be no movement, there is no force operating to homogenize these underlying parameters in such cases.

3.3   Differences due to hysteresis

Now consider two economies with identical ρ, δ values that lie in the hysteresis area between the bold lines in Figure 6. Country r is at a steady state like that sketched at point A, in Figure 5 while the poor country p is at a steady state corresponding to no skill acquisition – corresponding to the point B in Figure 5. In such a case, skilled individuals have a strict incentive to migrate to the place where skills are scarce.

Proposition 6. *When cross-country differences are driven by hysteresis and two countries with identical fundamentals have different levels of development, there is incentive for skilled individuals to migrate from the rich to the poor country.*

Here, the migration incentives have a purely neo-classical favour. The country abundant in skilled managers will see out-migration to the otherwise identical country which has not yet developed. Moreover, the introduction of a small number of skilled managers into the poor country will see it transition along the path in Figure 5 from the point B to eventually reach the same level of output per capita as the rich country, with full use of the modern technology. This "seeding" of skilled individuals and the eventual mushrooming of these skills throughout the sector, provides an interpretation of the Bangladeshi (garment) and Indonesian (lumber)

examples in the introduction. In both cases, small numbers of skilled workers were transplanted into the poor country – by training locals in the Bangladesi example, and by transplanting skilled Koreans in the Indonesian example. The diffusion of their skills through on the job learning to future entrepreneurs then lead to rapid and large follow-on effects.

4   Implications for Growth Accounting

A natural question to ask is whether the process of managerial skill acquisition we have presented could be an important component in development process in the sense of potentially explaining a substantial fraction of cross-country differences in income. At a first pass, the answer from the growth accounting literature appears to be a clear no. This literature (see for example Klenow and Rodriguez-Clare (1997)) finds that skills, whether acquired on-the-job or by schooling, account for only a small fraction of income differences across countries. In particular, in order to evaluate the potential role of on-the-job skill acquisition, this literature uses estimates of the returns to job experience to calculate the share of income paid to on-the-job acquired skills. This share of income is found to be small and it is generally inferred that no story of on-the-job skill acquisition can have the potential to explain an important fraction of cross-country income differentials. In this section, we argue that such an inference would not follow if the data were generated by the type of managerial skill acquisition process presented here. In this discussion, we will examine what our model implies about standard growth accounting exercises, under the assumption that returns on-the-job are estimated using experience profiles. We will illustrate why standard growth accounting exercises would not properly assess the role of on-the-job learning if the data were generated by our model. Since the impact upon growth accounting depends upon whether a cross-section of countries or a time-series is considered, we explore both cases.

4.1   Returns to experience in a cross-section

Growth accounting in a cross-section involves decomposing differences in (log) income-per-worker into components associated with factor accumulation and a residual. For example, 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 $\frac{EXP_i}{L_i}$ is experience per worker in country i, α_i is the share paid to experience, X_i is some composite of physical and human capital, (usually entering seperately, but bundled together for

simplicity here) together with its share, γ_i and A_i is the residual. The term α_i can be computed using estimates from a Mincer type earnings regression based on micro data. The residual term, A_i, is interpreted as country $i's$ productivity. In explaining cross-country differences in output per capita say between two countries i and j, the term:

$$\frac{\ln A_i - \ln A_j}{\ln \frac{Y_i}{L_i} - \ln \frac{Y_j}{L_j}}, \tag{22}$$

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

Now consider an accounting exercise aiming to explain level differences in income between countries i and j, using observations drawn from the steady state. Income per-capita is higher in country i because it has more managerial skills, which have been accumulated through the use of the new technology. However, returns to work experience estimated from micro data is precisely zero in country i. This is because managerial skills are abundant in the a steady state and hence have a zero return. Consequently, all of the difference in income levels that was caused by differences in on-the-job learning will be picked up by the residual. This may lead researchers to incorrectly interpret the cause of the difference in income per capita between i and j as being due to a disembodied factor, such as differences in technological opportunities, since it is attributed to the residual.

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

4.2 Returns to experience in a time series

The previous discussion illustrated why growth accounting exercises, when performed on a cross-section of countries, may greatly understate the role of on-the-job learning in explaining levels of

income-per-worker. In this section we examine whether such an inference problem will also arise when growth accounting is performed using time series data for a given country. Thus consider a single country that is experiencing growth in output per capita according to our process of managerial skills acquisition. In attempting to pick up the contribution of various factors to the country's growth over time, one would perform the following decomposition:

$$\frac{\dot{y}}{y} = \alpha_t \frac{\dot{\exp}}{\exp} + \gamma_t \frac{\dot{x}}{x} + \frac{\dot{A}}{A},$$

where x, y and exp are all in per-capita terms, and A represents the TFP residual. If experience were directly measured, i.e. if the utilized data included information about the technology with which workers work, then this procedure would provide an accurate depiction of experience's (or managerial skills') contribution to the growth in per capita output, as $\frac{\dot{\exp}}{\exp}$ would be positive while the economy grows and α_t would be greater than zero. It is thus, conceptually at least, possible to estimate the impact of all learning by doing on output per capita if the appropriate data were available.

However, direct measures of on-the-job skills are generally not available. Instead skills learned through experience are usually proxied as a weighted average of the economy's age structure, or a weighted average of time in the work-force. In our example, economy i experiences growth in output per capita all the way through its transition due to an increase in the fraction of the population which has managerial skill. But the age, and hence measured experience, is unchanged through the transition. Thus proxying experience by the age structure or time in the labor force would, in data generated by our model, lead to the finding that $\frac{\dot{\exp}}{\exp} = 0$. The outcome is therefore the same as in the cross-section: a growth accounting exercises performed on a time series will most likely attribute all of the on-the-job learning to the residual term, $\frac{\dot{A}}{A}$. This again would lead to the erroneous attribution of country $i's$ income growth to an unexplained residual factor instead of its true source; managerial skills purposefully accumulated on the job. We summarize this section with the following proposition:

Proposition 7. *The estimated returns to skills acquired on the job, reflected through standard growth accounting procedures, understates the true value of such skills.*

5 Discussion

We have assumed the existence of perfectly competitive labor markets which can adequately coordinate the acquisition of on-the-job managerial skills throughout. There are at least two

reasons to doubt this in reality: labor may be liquidity constrained; and firms may not be able to write binding long-term contracts with labor. In this section we briefly discuss these issues, leaving to future research a more thorough examination.

As we saw, in transition, there are skill premia, $\lambda > 0$. Unskilled labor have low marginal products and accept low wages in order to learn the managerial skills and receive higher wages in future. However, especially at the start of this process, the net present value of such skills is high. Thus competitive factor markets will price wages for the unskilled at low, and perhaps negative levels. We have not constrained wages for the unskilled downwards, but in reality it is extremely unlikely that, in developing countries, labor could pay for the future value of the organizational skills that they acquire today, through low or negative wages. Individuals are likely to be constrained by present consumption needs, and to be unable to finance present consumption out of future earnings.

This would be less of a problem if labor could write binding future employment contracts with employers (or outside financiers). With such contracts, labor could be paid a wage sufficient to maintain subsistence and above the market clearing wage while unskilled, and then, under contract, be forced to repay this by working as managers at wages below market clearing when they are ready to do so. The existence of such contracts would allow full mimicry of the competitive allocation that we have analyzed above. However, it is extremely unlikely that such contracts could work in reality. Firstly, in many countries, few courts would uphold the right of employers to restrict labor through contracts such as those above. Secondly, the discussion assumes away the possibility of worker moral-hazard, which is likely to become particularly severe when workers are being paid below market wages and would like to precipitate an employment termination. This problem could again be solved by contracts which allowed labor mobility but enforced repayment to either the firm or financier of the initial skills acquisition. However, even in developed countries, such financing schemes are far from complete.

If such on-the-job managerial skills acquisition does play a central role in productivity improvement, then it should be the case that countries that have experienced dramatic growth should have exhibited institutions that, to some extent, were able to overcome these problems. In contrast, the lack of such institutions may underpin the problems of skilling and technology acquisition that have been identified in many LDCs today.

Institutions that come to mind are those that somehow link labor to firms for long stretches

of time and thus provide incentives for firms to overcome workers' liquidity constraints and accumulate the necessary skills without worrying about future appropriability of returns. Or, institutions that allow workers to overcome these problems on their own, and thus become skilled. There is certainly evidence of such institutions in both Japan and South Korea which were both characterized by labor markets with significant worker bonding. There is also evidence of reduced labor mobility increasing firms' incentives to induce on-the-job skills acquisition in the developing past of industrialized countries, see for example Bessen (2003) and the case of 19th century Massachusetts cotton manufacturing. Singapore, in contrast, followed a different approach which explicitly recognized the externality to such on the job training in a market where worker mobility was assured. The government there set up a Skill Development Fund which collected a levy of one percent of payroll from employers to subsidize training of low paid workers (Lall 1999, p. 36). Alternatively, Germany developed a 'cooperative' training system which lead to practical training being coordinated between employer, employees and government. Apprentices took low wages, employers contributed to part of the costs, and government provided some of the infrastructure, while union groups took an active role in discouraging worker poaching (Lall 1999, p.38). A detailed analysis of these country studies is beyond the scope of this paper, but further examination of such cases is certainly motivated by the findings here.

6   Conclusion

In this paper we have presented a model of managerial skill acquisition where unskilled individuals invest to obtain skills by accepting employment in firms where managerial skills are present. This process of skills acquisition is related to, but distinct from, a previously large literature on skills acquisition through learning-by-doing. Learning-by-doing emphasizes increases in productivity arising from repetition of tasks through time. Our process of managerial skills acquisition involves learning from working with a manager, and thus arises when undertaking distinct and complementary tasks. This process may have potentially important implications for understanding why technologies may be so difficult to transplant across countries. Unlike the learning-by-doing literature, and previous studies which have similarly emphasized the role of emodied skills (as opposed to disembodied knowledge), the process we posit here yields implications which we demonstrate to be consistent with the empirical record.

Although this process of managerial skill acquisition is modelled in a manner that allows

knowledge to be entirely internalized by individuals, we have shown that it nevertheless leads to outcomes that resemble those obtained in endogenous growth models with market imperfections. For example, we show that in this setup, agents will purposefully accumulate managerial skill knowing full well that the return to these skills will dissipate over time and, in the steady state, possibly be equal to zero. We have shown how such a process could lead astray growth accounting exercises since differences across countries in managerial skill would most likely be attributed to differences in total factor productivity as opposed to being properly attributed to on-the-job experience. Finally, we have also shown how this model can explain why managers may not migrate from rich countries, where they are abundant, to poor countries where there are scarce, even though the underlying technology satisfies constant returns to scale. The explanation of this latter effect is the presence of a teaching premium for managers in the rich country which will generally be higher that in the poor country.

There are many possible extensions to this model. As previously discussed, it may be warranted to examine the properties of this type of on-the-job learning process in the absence of perfect markets. This could help understand the institutional and policy features of less developed economies that are conducive (or inimical) to this process of skills acquisition when markets do not fully reward it. An alternative type of market failure may arise when the modern sector is subject to thick market externalites. In the presence of even minor externalities of this sort, the process of managerial skill acquisition as modelled here could explain the emergence of a poverty trap whereby a poor country does not invest in managerial skill acquisition simply because the modern sector is small due to the lack of managers.

References

[1] Acemoglu, D. P. Aghion and F. Zilibotti (2003) Vertical Integration and Distance to Frontier, forthcoming, *Journal of the European Economic Association*

[2] Acemoglu, D. and F. Zilibotti (2001) Productivity Differences, *Quarterly Journal of Economics*, vol. 115(3), pp.563-606.

[3] Acemoglu, D. and A. Newman (2002) The Labor Market and Corporate Structure, *European Economic Review*, volume 46, pp. 1733-1756.

[4] Aghion, P. and P. Howitt (1992) Aghion and Howitt (1992) "A Model of Growth Through Creative Destruction" *Econometrica*, 60, 323-351.

[5] Alchian, A. (1963): Reliability of progress curves in airframe production. *Econometrica*, 31(4):679-693.

[6] Arrow, K. J., (1962a) Economic welfare and the allocation of re-sources for inventions. In Richard R. Nelson, editor, The Rate and Direction of Inventive Activity. Princeton University Press and NBER.

[7] Arrow, Kenneth J (1962b)., The Economic Implications of Learning by Doing, *Review of Economic Studies,* June 1962, Vol. 29, No. 3, pp. 155-173

[8] Barro R.J. and X. Sala-i-Martin (1995), *Economic growth*, McCraw-Hill, New York.

[9] Becker, G.S. (1975) Human capital: A theoretical and empirical analysis, with special reference to education. (2nd ed.). National Bureau of Economic Research. New York, NY: Columbia University Press.

[10] Benhabib, J. and M. Spiegel (2003) Human capital and technology diffusion, Federal Reserve Bank of San Francisco, Development Research Institute working paper # 2003-03.

[11] Bessen, J. (2003) Technology and Learning by Factory Workers: The Stretch-Out at Lowell, 1842, Journal of Economic History; 63(1): 33-64

[12] Biggs,-Tyler; Shah,-Manju; Srivastava,-Pradeep (1995) Technological capabilities and learning in African enterprises, Technical Paper, no. 288. Africa Technical Department Series. Washington, D.C.: World Bank.

[13] Chari, V.V. and H. Hopenhayn (1991) Vintage human capital, growth and the diffusion of new technology, Journal of Political Economy, 99, 61 1142-1165.

[14] Devarajan, S., W. Easterly, and H. Pack) Low Investment is not the Constraint on African Development Economic Development and Cultural Change, April 2003, Volume 51, No. 3.

[15] Easterly, W. (2002) The elusive quest for growth: economists adventures and misadventures in the tropics, Cambridge, MIT press.

[16] Enos, M. D., Kehrhahn, M. T. & Bell, A. (2003). Informal learning and the transfer of learning: How managers develop proficiency. Human Resource Development Quarterly, 14(4), 369-387.

[17] Fafchamps M. and M. Soderbom (2004) Wages and labor management in African manufacturing, mimeo, Oxford University.

[18] Feldman, M. (2001) The entrepreneurial event revisited: firm formation in a regional context, Industrial and Corporate Change, 10 (4) 861-891.

[19] Grossman, G. and E. Helpman (1991), Innovation and Growth in the Global Economy, Cambridge: MIT Press

[20] Irwin, D.A and P. Klenow (1994) Learning by doing spillovers in the semi-conductor industry, Journal of Political Economy, 102(6) 1200-1227.

[21] Keller, W. (2004) International Technology Diffusion, Journal of Economic Literature, 42(3) 752-82.

[22] Klenow, P. and A. Rodriguez-Clare (1997) The Neoclassical Revival in Growth Economics: Has It Gone Too Far? in Bernanke,-Ben-S.; Rotemberg,-Julio-J., eds. NBER macroeconomics annual 1997. Cambridge and London: MIT Press, 1997; 73-103.

[23] Lall, S. (1999) Competing with labour: skills and technology in developing countries ILO, Issues in Development working paper # 31.

[24] Lester, R., and M. Piore (2004) Innovation: The Missing Dimension. Cambridge, MA. Harvard University Press.

[25] Lucas, R. (1993) Making a Miracle, Econometrica, March 61(2): 251-72.

[26] Maddison, A. (1995) Monitoring the world economy: 1820-1992, Paris, OECD.

[27] Mankiw, N. G. (1995) The growth of nations, Brookings Papers on Economic Activity, 275-310.

[28] McCall, M., Lombardo, M., & Morrison, A. (1988). The lessons of experience: How successful executives develop on the job. Lexington, MA: Lexington Press.

[29] Parente, S. and E. Prescott (2002) Barriers to Riches, Cambridge, MIT Press.

[30] Rapping, L (1965): "Learning and World War II Production Functions." Review of Economics and Statistics, 47(1):81-86.

[31] Rhee, Y. and T. Belot (1990) Export Catalysts in Low-Income Countries: A Review of Eleven Success Stories, World Bank Discussion Paper number 72.

[32] Romer, P. (1995) The Growth of Nations: Comment, Brookings Papers on Economic Activity 0(1): 313-20.

[33] Saxenian, A. (1994) Regional advantage: culture and competition in Silicon Valley and Route 128, Harvard University Press, Cambridge, MA.

[34] Searle, Allan D. (1945): "Productivity Changes in Selected Wartime Shipbuilding Programs." Monthly Labor Review, (December):1132-1147.

[35] Shell, K. (1966) Toward a theory of inventive activity and capital accumulation, American Economic Review,

[36] Stokey, N. (1988) Learning by Doing and the Introduction of New Goods, Journal of Political Economy. August 96(4): 701-17.

[37] Young, A. (1991) Learning by Doing and the Dynamic Effects of International Trade Quarterly Journal of Economics; 106(2): 369-405

7   Appendix

7.0.1   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

$$
\begin{aligned}
\dot{\lambda} &= F_2\left(S, (1-S)\right) - F_1\left(S, (1-S)\right) \\
&\quad + \lambda\left[(\delta + \rho) + \Omega\left(S, (1-S)\right) + \left\{\Omega_2\left(S, (1-S)\right) - \Omega_1\left(S, (1-S)\right)\right\}(1-S)\right]
\end{aligned} \quad (23)
$$

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 A1. Beyond the S' locus, $\gamma < 1$, and λ must take value 0. The directional arrows indicate the movements of λ .
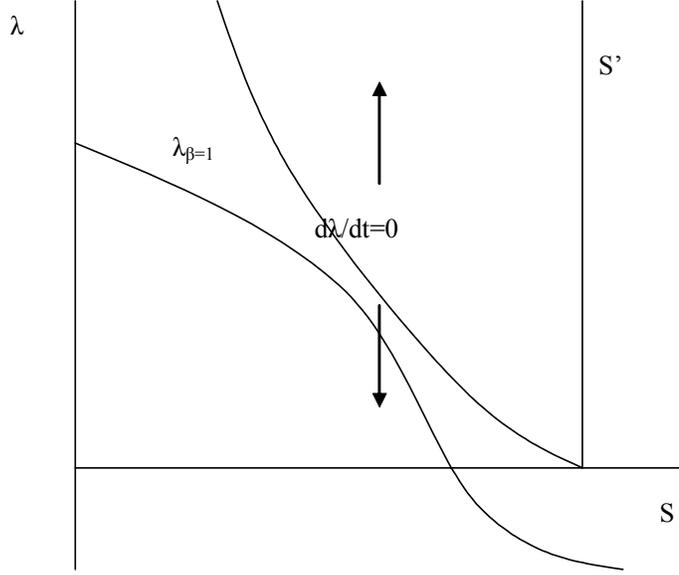
36

Figure A1

When crossing points exist, the system moves into the $\beta < 1$ range, the dynamics of which we now analyze.

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.

$$\frac{\overline{w} - F_2\left(S, (1-S)\right)}{\Omega + \Omega_2\left(S, 1-S\right)\left(1-S\right)} = \frac{F_1\left(S, 1-S\right) - F_2\left(S, 1-S\right)}{\delta + \rho + \Omega\left(S, (1-S)\right) + \left[\Omega_2\left(S, (1-S)\right) - \Omega_1\left(S, (1-S)\right)\right]\left(1-S\right)}. \tag{24}$$

There may be multiple roots to this equation.

In the $\beta < 1$ region, the value of β is implicitly defined by the condition:

$$\overline{w} - F_2\left(S_t, \beta_t\left(1-S_t\right)\right) = \lambda_t\left[\Omega\left(S_t, \beta_t\left(1-S_t\right)\right) + \Omega_2\left(S_t, \beta_t\left(1-S_t\right)\right)\beta_t\left(1-S_t\right)\right] \tag{25}$$

and we have:

$$\dot{\lambda} = \left(\rho + \delta\right)\lambda + \overline{w} - F_1\left(S, \beta\left(1-S\right)\right) - \lambda\beta\left(1-S\right)\left(\Omega_1\left(S, \beta\left(1-S\right)\right)\right)$$

The $\dot{\lambda} = 0$ locus is defined for $\beta < 1$ only if there exist points of intersection defined by equation (24). Denote the set of all S at such intersection points by Λ. For values of S, λ below the locus $\lambda_{\beta=1}$ the $\dot{\lambda} = 0$ line is given by:

$$\lambda = \frac{F_1\left(S, \beta\left(1-S\right)\right) - \overline{w}}{\rho + \delta - \Omega_1\left(S, \beta\left(1-S\right)\right)\beta\left(1-S\right)}. \tag{26}$$

It follows from homotheticity that such locuses are horizontal lines

37

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

Proof: Rewrite expression (26) as $\lambda = \frac{F_1(S, \beta(1-S)) - \overline{w}}{\rho + \delta - \widetilde{\Omega}_1(S, \beta(1-S))}$ and differentiate with respect to S yielding:

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

$$(27)$$

Differentiating the first order condition yields:

$$-\lambda \left(\widetilde{\Omega}_{21} + \widetilde{\Omega}_{22} \left(\frac{d\beta}{dS}(1-S) - \beta \right) \right) - \frac{d\lambda}{dS} \widetilde{\Omega}_2 = F_{21} + \left(\frac{d\beta}{dS}(1-S) - \beta \right) F_{22}$$

$$\Rightarrow \left(\frac{d\beta}{dS}(1-S) - \beta \right) = -\frac{F_{21} + \frac{d\lambda}{dS}\widetilde{\Omega}_2 + \lambda\widetilde{\Omega}_{21}}{F_{22} + \widetilde{\Omega}_{22}\lambda}.$$

Substituting into (27) and rearranging yields:

$$\frac{d\lambda}{dS} \left(1 + \frac{\widetilde{\Omega}_2}{F_{22} + \widetilde{\Omega}_{22}\lambda} \left(F_{12} + \widetilde{\Omega}_{12} \right) \right) = \left(F_{11} - \frac{F_{21} + \lambda\widetilde{\Omega}_{21}}{F_{22} + \lambda\widetilde{\Omega}_{22}} F_{12} \right) \left(\rho + \delta + \widetilde{\Omega}_1 \right) \quad (28)$$

$$+ (F_1 - \overline{w}) \left(\widetilde{\Omega}_{11} - \frac{F_{21} + \lambda\widetilde{\Omega}_{21}}{F_{22} + \lambda\widetilde{\Omega}_{22}} \widetilde{\Omega}_{12} \right). \quad (29)$$

Since $\widetilde{\Omega}$ and F are HD1, we have:

$$F_{11}S + F_{12}\beta(1-S) = 0$$

$$F_{21}S + F_{22}\beta(1-S) = 0$$

$$\widetilde{\Omega}_{11}S + \widetilde{\Omega}_{12}\beta(1-S) = 0$$

$$\widetilde{\Omega}_{21}S + \widetilde{\Omega}_{22}\beta(1-S) = 0.$$

These equalities imply that the terms in large parentheses on the right hand side of (29) equal zero, so that $\frac{d\lambda}{dS} = 0.\blacksquare$

A situation with two crossing points is depicted in Figure A2. Once again, the arrows depict the dynamics of λ
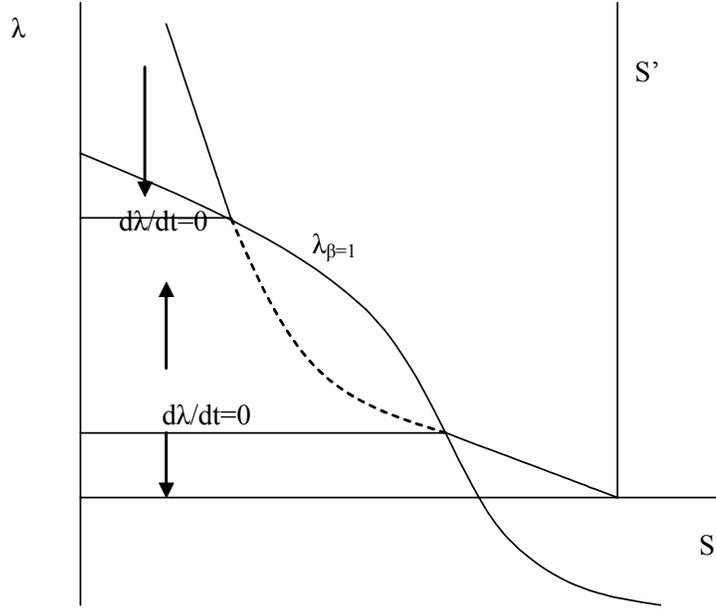
Figure A2

7.0.2   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

$$\Omega\left(S, (1 - S)\right)(1 - S) = \delta S, \tag{30}$$

denoted $\overline{S} \in (0, 1)$. It follows that the locus is bounded away from 1 because $\Omega\left(S, 0\right)(0) = 0$. In this case, the $\dot{S} = 0$ equation corresponds to a vertical line at $\overline{S}$, as depicted in Figure A3. Note that $\overline{S}$ can be either to the right or left of S'.

Figure A3

It is also possible for there to be an intersection between $\dot{S} = 0$ and $\lambda_{\beta=1}$, which is defined by the pair:

$$S = \overline{S}, \lambda = \frac{\overline{w} - F_2\left(\overline{S},\left(1 - \overline{S}\right)\right)}{\Omega + \Omega_2\left(\overline{S}, 1 - \overline{S}\right)\left(1 - \overline{S}\right)}.$$
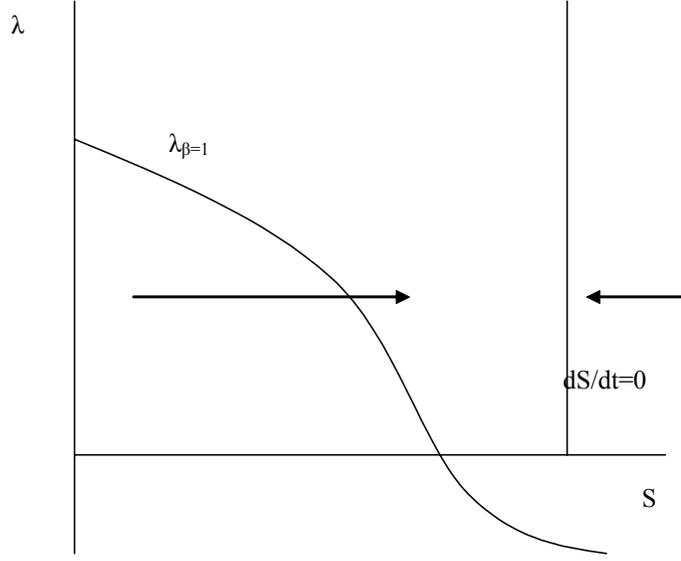
When such a crossing does not exist in the positive orthant, i.e. when $\overline{w} - \theta F_2\left(\overline{S}, 1 - \overline{S}\right) < 0$, the movement of the system is always towards $S = \overline{S}$ as depicted in Figure A3. Conversely, if a crossing exists, it is unique, and then the following 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 $\frac{d\lambda}{dS} = 0$.*

Proof: $\dot{S} = 0$ implies

$$\Omega\left(S, \beta\left(1 - S\right)\right)\beta\left(1 - S\right) \equiv \widetilde{\Omega}\left(S, \beta\left(1 - S\right)\right) = \delta S. \tag{31}$$

Differentiating yields

$$\widetilde{\Omega}_1 + \widetilde{\Omega}_2\left(\frac{d\beta}{dS}\left(1 - S\right) - \beta\right) = \delta.$$

As in the previous Lemma, differentiating the first order condition yields:

$$\left(\frac{d\beta}{dS}\left(1 - S\right) - \beta\right) = -\frac{F_{21} + \frac{d\lambda}{dS}\widetilde{\Omega}_2 + \lambda\widetilde{\Omega}_{21}}{F_{22} + \widetilde{\Omega}_{22}\lambda}.$$

Combining these two expressions leads to:

$$sign\left(\frac{d\lambda}{dS}\right) = sign\left(\left(\delta - \widetilde{\Omega}_1\right)\left(F_{22} + \widetilde{\Omega}_{22}\lambda\right) + \left(F_{21} + \lambda\widetilde{\Omega}_{21}\right)\widetilde{\Omega}_2\right). \tag{32}$$

40

Since $\widetilde{\Omega}$ is HD1, $\widetilde{\Omega}_1 S + \widetilde{\Omega}_2 \beta (1 - S) = \widetilde{\Omega} = \delta S$, where the last inequality holds at $\dot{S} = 0$. Using this expression to substitute for $\widetilde{\Omega}_1$ in (32) yields:

$$sign\left(\frac{d\lambda}{dS}\right) = sign\left(\left(\widetilde{\Omega}_2 \beta \frac{1-S}{S}\right)\left(F_{22} + \widetilde{\Omega}_{22}\lambda\right) + \left(F_{21} + \lambda\widetilde{\Omega}_{21}\right)\widetilde{\Omega}_2\right). \tag{33}$$

Since F_i and $\widetilde{\Omega}_i$ are HD0, we have:

$$SF_{21} + \beta(1 - S)F_{22} = 0$$
$$S\widetilde{\Omega}_{21} + \beta(1 - S)\widetilde{\Omega}_{22} = 0.$$

Using these to substitute into (33) yields

$$sign\left(\frac{d\lambda}{dS}\right) = sign\left(\left(\widetilde{\Omega}_2 \beta \frac{1-S}{S}\right)\left(F_{22} + \widetilde{\Omega}_{22}\lambda\right) - \left(\beta\frac{1-S}{S}F_{22} + \lambda\beta\frac{1-S}{S}\widetilde{\Omega}_{22}\right)\widetilde{\Omega}_2\right) = 0.$$

∎

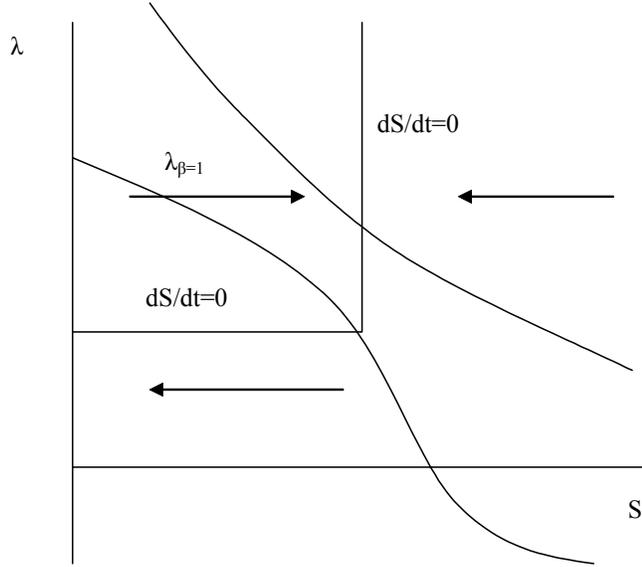Figure A4 depicts the dynamics for S when such a point of intersection exists.

Figure A4

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

Proof:

$$\frac{d\lambda_{\beta=1}}{dS} = sign[-\left(\Omega - \Omega_2\left(1 - S\right)\right)\left(F_{21} - F_{22}\right) - \left(\overline{w} - F_2\right)\left(\Omega_1 - \Omega_2 + \left(\Omega_{21} - \Omega_{22}\right)\left(1 - S\right) - \Omega_2\right)]$$

The first term in the square brackets is negative. From HDO of Ω we have:

$$\Omega_1 S + \Omega_2\left(1 - S\right) = 0 \tag{34}$$

$$\Rightarrow \left(\Omega_{21} - \Omega_{22}\right)\left(1 - S\right) + \left(\Omega_1 - \Omega_2\right) = \left(\Omega_{12} - \Omega_{11}\right) S$$

which implies that the second term becomes

$$-(-\Omega_2 + \left(\Omega_{12} - \Omega_{11}\right) S) \tag{35}$$

From the homogeneity of degree -1 of the function Ω_1 we have

$$\Omega_{11} S + \Omega_{12}\left(1 - S\right) = -\Omega_1$$

substituting from (34) 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 term, expression (35), is negative so that the whole expression is negative.■

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

Proof: $\dot{\lambda} = 0|_{\beta=1} \Rightarrow$

$$\lambda = \frac{F_1 - F_2}{\delta + \rho + \widetilde{\Omega}_2 - \widetilde{\Omega}_1}$$

$$and$$

$$(sign)\frac{d\lambda}{dS} = sign\left[\begin{array}{c} \left(F_{11} - F_{12} - F_{21} + F_{22}\right)\left(\delta + \rho + \widetilde{\Omega}_2 - \widetilde{\Omega}_1\right) \\ -\left(F_1 - F_2\right)\left(\widetilde{\Omega}_{21} - \widetilde{\Omega}_{22} - \widetilde{\Omega}_{12} + \widetilde{\Omega}_{12}\right) \end{array}\right]. \tag{36}$$

Since $\widetilde{\Omega}_{ij} > 0$, $\widetilde{\Omega}_{ii} < 0$, $F_{ii} < 0$ and $F_{ij} > 0$ it follows that $(sign)\frac{d\lambda}{dS} < 0$.■

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 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 the case depicted in the figure below.

42
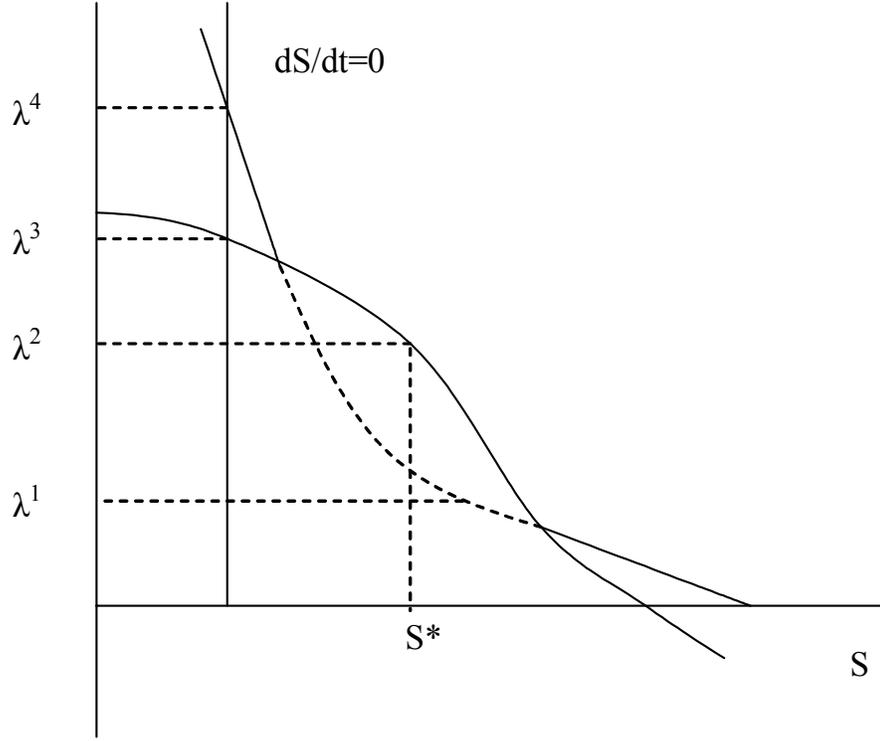
Figure A5

For this proof it is simpler to use the $\widetilde{\Omega}$ function :

$$\widetilde{\Omega}\left(S, 1-S\right) \equiv \Omega\left(S, 1-S\right)\left(1-S\right)$$

Since Ω is HDO the function $\widetilde{\Omega}$ is HD1.

Assign the following λ values: For $\overline{S}$ defined in (30) we have:

$$\lambda_4 = \frac{F_1 - F_2}{\delta + \rho + \widetilde{\Omega}_2 - \widetilde{\Omega}_1}$$

$$\lambda_3 = \frac{\overline{w} - F_2}{\widetilde{\Omega}_2}$$

Consider any two elements S_1, S_2 of Λ: and consider an S^* with $S_1 < S^* < S_2$, such that the $\dot{\lambda} = 0$ function defined for $\beta = 1$ lies below the $\lambda_{\beta=1}$ line. Then define

$$\lambda_2 = \frac{\overline{w} - F_2^*}{\widetilde{\Omega}_2^*}$$

$$\lambda_1 = \frac{F_1^* - F_2^*}{\delta + \rho + \widetilde{\Omega}_2^* - \widetilde{\Omega}_1^*}$$

$$where \; F^* \equiv F\left(S^*, 1-S^*\right) \; and \; \widetilde{\Omega}^* = \widetilde{\Omega}\left(S^*, 1-S^*\right).$$

43

Clearly, for this case to exist it is necessary that: $\lambda_4 > \lambda_3 > \lambda_2 > \lambda_1$.

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

$$\widetilde{\Omega}_2 = \left(\delta - \widetilde{\Omega}_1\right) \frac{\overline{S}}{1 - \overline{S}}. \tag{37}$$

For the remainder of the proof $\widetilde{\Omega}$ and F will denote the respective functions evaluated at $S = \overline{S}$. Applying these, $\lambda_4 > \lambda_3$ implies that:

$$\frac{F_1 - F_2}{\delta + \rho + \widetilde{\Omega}_2 - \widetilde{\Omega}_1} > \frac{\overline{w} - F_2}{\widetilde{\Omega}_2}$$

$$\Rightarrow \quad \frac{F_1 - F_2}{\delta + \rho + \left(\delta - \widetilde{\Omega}_1\right)\frac{\overline{S}}{1-\overline{S}} - \widetilde{\Omega}_1} > \frac{\overline{w} - F_2}{\left(\delta - \widetilde{\Omega}_1\right)\frac{\overline{S}}{1-\overline{S}}}$$

$$\Rightarrow \quad \frac{F_1 - F_2}{\rho + \left(\delta - \widetilde{\Omega}_1\right)\frac{1}{1-\overline{S}}} > \frac{\overline{w} - F_2}{\left(\delta - \widetilde{\Omega}_1\right)\frac{\overline{S}}{1-\overline{S}}}$$

$$\Rightarrow \quad \frac{F_1 - F_2}{(1 - \overline{S})\rho + \left(\delta - \widetilde{\Omega}_1\right)} > \frac{\overline{w} - F_2}{\left(\delta - \widetilde{\Omega}_1\right)\overline{S}}$$

$$\Rightarrow \quad \left(\delta - \widetilde{\Omega}_1\right)\overline{S} F_1 - \left(\delta - \widetilde{\Omega}_1\right)\overline{S} F_2 > (1 - \overline{S})\rho(\overline{w} - F_2) + \left(\delta - \widetilde{\Omega}_1\right)(\overline{w} - F_2)$$

$$\Rightarrow \quad \left(\delta - \widetilde{\Omega}_1\right)\overline{S} F_1 - \left(\delta - \widetilde{\Omega}_1\right)\overline{S} F_2 > (1 - \overline{S})\rho(\overline{w} - F_2) + \overline{S}\left(\delta - \widetilde{\Omega}_1\right)(\overline{w} - F_2)$$
$$+ (1 - \overline{S})\left(\delta - \widetilde{\Omega}_1\right)(\overline{w} - F_2) \tag{38}$$

$$\Rightarrow \quad \left(\delta - \widetilde{\Omega}_1\right)\overline{S}(F_1 - \overline{w}) > (1 - \overline{S})\left(\delta - \widetilde{\Omega}_1 + \rho\right)(\overline{w} - F_2)$$

$$\Rightarrow \quad \frac{F_1 - \overline{w}}{\delta - \widetilde{\Omega}_1 + \rho} > \frac{\overline{w} - F_2}{\left(\delta - \widetilde{\Omega}_1\right)\frac{\overline{S}}{(1-\overline{S})}}$$

$$\Rightarrow \quad \frac{F_1 - \overline{w}}{\delta - \widetilde{\Omega}_1 + \rho} > \frac{\overline{w} - F_2}{\widetilde{\Omega}_2}. \tag{39}$$

Now notice that

$$\frac{F_1^* - F_2^*}{\delta + \rho + \widetilde{\Omega}_2^* - \widetilde{\Omega}_1^*} > \frac{F_1^* - F_2^*}{\delta + \rho - \widetilde{\Omega}_1^*} > \frac{F_1^* - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1^*} > \frac{F_1^* - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(\overline{S})}. \tag{40}$$

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

$$\frac{\overline{w} - F_2^*}{\widetilde{\Omega}_2^*} < \frac{\overline{w} - F_2^*}{\widetilde{\Omega}_2(\overline{S})}, \tag{41}$$

since $\widetilde{\Omega}_2^* > \widetilde{\Omega}_2(\overline{S})$. Using (40) and (41) it then follows that $\lambda_2 > \lambda_1$ implies:

$$\frac{\overline{w} - F_2^*}{\widetilde{\Omega}_2(\overline{S})} > \frac{F_1^* - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(\overline{S})}. \tag{42}$$

44

Once again utilizing (37) inequalities (39) and (42) can be expressed respectively as:

$$\frac{(F_1 - \overline{w})\,\overline{S}}{\delta - \widetilde{\Omega}_1 + \rho} > \frac{(\overline{w} - F_2)\,(1 - \overline{S})}{\delta - \widetilde{\Omega}_1} \tag{43}$$

$$\frac{(\overline{w} - F_2^*)\,(1 - \overline{S})}{\delta - \widetilde{\Omega}_1} > \frac{(F_1^* - \overline{w})\,\overline{S}}{\delta + \rho - \widetilde{\Omega}_1\,(\overline{S})}. \tag{44}$$

Adding the two inequalities above makes the $\overline{w}$ terms cancels and yields:

$$\frac{F_1 \overline{S}}{\delta - \widetilde{\Omega}_1 + \rho} - \frac{F_2^*\,(1 - \overline{S})}{\delta - \widetilde{\Omega}_1} > \frac{\overline{S} F_1^*}{\delta + \rho - \widetilde{\Omega}_1\,(\overline{S})} - \frac{F_2\,(1 - \overline{S})}{\delta - \widetilde{\Omega}_1}.$$

Rearranging terms in this expression implies:

$$F_1 \overline{S}\left(\delta - \widetilde{\Omega}_1\right) - F_2^*\,(1 - \overline{S})\left(\delta - \widetilde{\Omega}_1 + \rho\right) > \overline{S} F_1^*\left(\delta - \widetilde{\Omega}_1\right) - F_2\,(1 - \overline{S})\left(\delta + \rho - \widetilde{\Omega}_1\,(\overline{S})\right)$$

$$\Rightarrow \quad F_1 \overline{S} + F_2\,(1 - \overline{S}) > \overline{S} F_1^* + F_2^*\,(1 - \overline{S}).$$

But since $S^* > \overline{S}$ 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 3) 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 $S = 0$, $\overline{w} - F_2\,(0,0) = \overline{w} > \lambda_t \Omega$, so that $\beta_t = 0$. But if $\beta_t = 0$ then $\dot{S} = 0$ at $S = 0$, since S is constrained to being 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 4.

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 $\overline{w} > \lambda_t \Omega$, so that $\dot{S} = 0$ since S is constrained to being non-negative. A steady state thus exists at the intersection of the $\dot{\lambda} = 0$ locus and the vertical axis, as depicted in Figure 5, point B. But another steady state exists in the $\beta = 1$ region at the intersection of the two locii.∎

Proof of Proposition 3:

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,

$$\frac{\overline{w} - F_2(\cdot)}{\Omega(\cdot) + \Omega_2(\cdot)(1-S)} - \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + [\Omega_2(\cdot) - \Omega_1(\cdot)](1-S)} > 0, \tag{45}$$

i.e. if the set Λ is non-empty. Suppose that $\rho + \delta = 0$. Then, necessarily, condition (45) fails. Because $F_1 > \overline{w}$ 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

$$\frac{\overline{w} - F_2(\cdot)}{\Omega(\cdot) + \Omega_2(\cdot)(1-S)} - \frac{F_1(\cdot) - F_2(\cdot)}{\delta + \rho + \Omega(\cdot) + [\Omega_2(\cdot) - \Omega_1(\cdot)](1-S)} = 0. \tag{46}$$

Let k^* denote the value of $\rho + \delta$ solving (46) with equality. If and only if $\rho + \delta < k^*$ the functions do not cross and the situation in Figure 3 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 (2). Consider $\overline{S}$ defined in (30). If and only if $\overline{S} > S'$, 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 3 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 $\delta^* = \frac{\Omega(S', 1-S')(1-S')}{S'}$.

Proof of part c: δ^{**} solves θ

$$F_2\left(\overline{S}, 1 - \overline{S}\right) = \overline{w}$$
$$\delta^{**} = \frac{\Omega\left(\overline{S}, \left(1-\overline{S}\right)\right)\left(1-\overline{S}\right)}{\overline{S}}$$

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 3. If $\delta + \rho \geq k^*$, there is an interesection and the case of Figure 5 applies, implying hysteresis. For $\delta \geq \delta^{**}$, the $\dot{S} = 0$ locus intersects the $\lambda_{\beta=1}$ line. Note that, from condition (3) $\delta^{**} > \delta^*$. 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:

$$\frac{\overline{w} - F_2\left(\overline{S}, 1-\overline{S}\right)}{\Omega\left(\overline{S}, 1-\overline{S}\right) + \Omega_2\left(\overline{S}, 1-\overline{S}\right)\left(1-\overline{S}\right)} = \frac{F_1\left(\overline{S}, 1-\overline{S}\right) - F_2\left(\overline{S}, 1-\overline{S}\right)}{\delta + \rho + \Omega\left(\overline{S}, 1-\overline{S}\right) + \left[\Omega_2\left(\overline{S}, 1-\overline{S}\right) - \Omega_1\left(\overline{S}, 1-\overline{S}\right)\right]\left(1-\overline{S}\right)}$$
$$\implies \rho^* = \frac{F_1(\cdot)\left[\Omega(\cdot) + \Omega_2(\cdot)\left(1-\overline{S}\right)\right] + F_2(\cdot)\left[\delta - \Omega_1\left(1-\overline{S}\right)\right] - \overline{w}\left[\Omega(\cdot) + \left(\Omega_2(\cdot) - \Omega_1\right)\left(1-\overline{S}\right) + \delta\right]}{\overline{w} - F_2(\cdot)}.$$

Note that since $\alpha F_1 + (1-\alpha) F_2 > \overline{w}$ 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 4 occurs, and the unique steady state is full use of the traditional technology. Under the converse, the situation in 5 occurs and hysteresis is the outcome.∎

Proof of Proposition 4:

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 parabus. From Proposition 3 recall that the steady state allocation of labor varies with the sum $\rho + \delta$ depicted by the downward sloping straight line in Figure 6. For given $\delta < \delta'$, there thus exists a unique value of ρ, denote it $\rho^c \equiv k^* - \delta$, such that for $\rho > \rho^c$, $\beta = 0$ (corresponding to the configuration in Figure 4) 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 $\rho > \rho^c$,

$$w^s(\rho) = F_1(Z, 1) + \lambda \widetilde{\Omega}_1(Z, 1), \tag{47}$$

where $Z \equiv \frac{\beta S}{1 - S}$. The value of λ is then given by:

$$\lambda = \frac{F_1(Z, 1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z, 1)}.$$

And Z is the smallest root of:

$$\frac{F_1(Z, 1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z, 1)} = \frac{\overline{w} - F_2(Z, 1)}{\widetilde{\Omega}_2(Z, 1)},$$

corresponding to the value of Z at the point A in Figure 4, or the point λ^p in Figure A6 below. By inspection of the figure it is immediate that, as ρ increases, the curved schedule shifts downward to the left, implying that λ falls. This implies that Z increases since λ also satisfies $\lambda = \frac{\overline{w} - F_2(Z,1)}{\overline{\Omega}_2(Z,1)}$. Since Z increases and λ falls, it is immediate from (47) 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, $\overline{w}$. 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 avaiable to them, i.e. $\beta = 1$. Consequently, returns are higher to the unskilled in the rich country. ∎
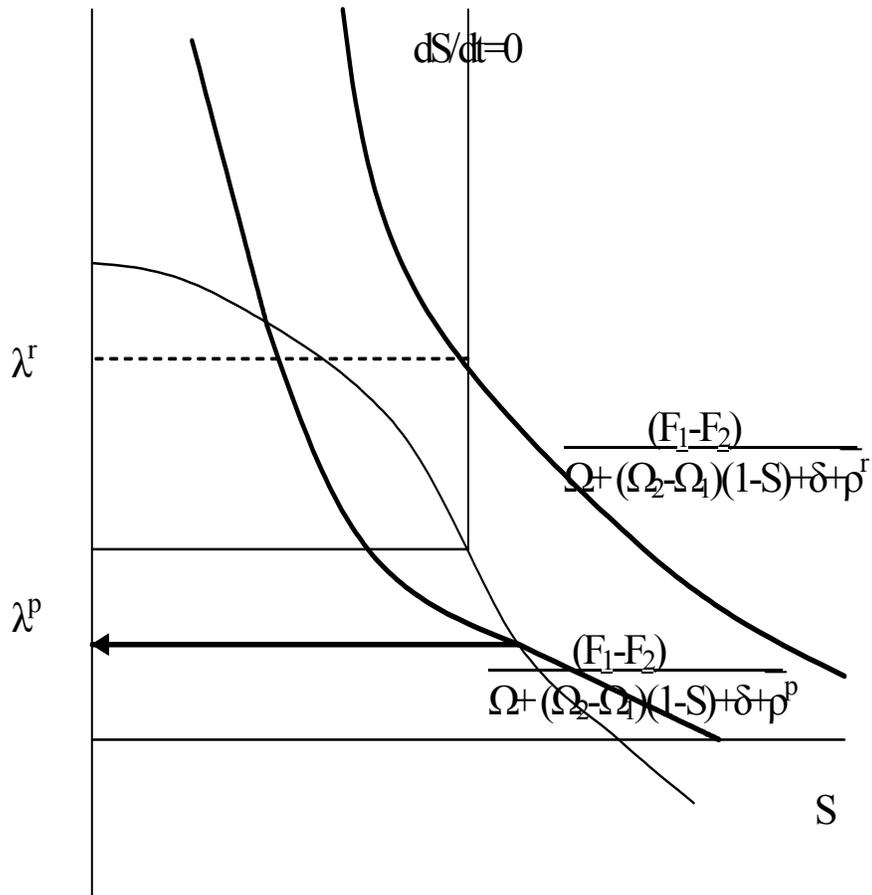

Figure A6

Proof of Proposition 5: At steady state A all individuals work in the modern sector

48

therefore:

$$w^s(A) = F_1\left(S^A, 1 - S^A\right) + \lambda \Omega_1\left(S^A, 1 - S^A\right)\left(1 - S^A\right).$$

Where steady state S is determined by

$$\delta S^A = \Omega\left(S^A, 1 - S^A\right)\left(1 - S^A\right)$$

$$and$$

$$\lambda = \frac{F_1\left(\left(S^A, 1 - S^A\right)\right) - F_2\left(\left(S^A, 1 - S^A\right)\right)}{\delta + \rho + \Omega\left(\left(S^A, 1 - S^A\right)\right) + \left[\Omega_2\left(\left(S^A, 1 - S^A\right)\right) - \Omega_1\left(\left(S^A, 1 - S^A\right)\right)\right]\left(1 - S^A\right)} \equiv \lambda^A.$$

This implies:

$$
\begin{aligned}
w^s(A) &= F_1\left(\left(S^A, 1 - S^A\right)\right) + \lambda^A \Omega_1\left(\left(S^A, 1 - S^A\right)\right)\left(1 - S^A\right) \\
&\equiv F_1\left(\left(S^A, 1 - S^A\right)\right) + \lambda^A \widetilde{\Omega}_1\left(\left(S^A, 1 - S^A\right)\right).
\end{aligned}
\tag{48}
$$

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

$$w^s(B) = F_1(Z, 1) + \lambda^B \widetilde{\Omega}_1(Z, 1),\tag{49}$$

where $Z \equiv \frac{\beta S}{1 - S}$. The value of λ is then given by:

$$\lambda = \frac{F_1(Z, 1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z, 1)}.$$

And Z is the smallest root of:

$$\frac{F_1(Z, 1) - \overline{w}}{\delta + \rho - \widetilde{\Omega}_1(Z, 1)} = \frac{\overline{w} - F_2(Z, 1)}{\widetilde{\Omega}_2(Z, 1)}.$$

At S^A since $\beta = 1$ we know that:

$$\overline{w} < F_2\left(\left(S^A, 1 - S^A\right)\right) + \lambda^A \widetilde{\Omega}_2\left(\left(S^A, 1 - S^A\right)\right)\tag{50}$$

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

$$F_2(Z, 1) + \lambda^B \widetilde{\Omega}_2(Z, 1) = \overline{w}.\tag{51}$$

But since by inspection of Figure 5 it is immediate that $\lambda^B > \lambda^A$ it follows from (50) and (51) that

$$F_2(Z, 1) + \widetilde{\Omega}_2(Z, 1) < F_2\left(S^A, 1 - S^A\right) + \widetilde{\Omega}_2\left(S^A, 1 - S^A\right).$$

This implies that $Z < \frac{S^A}{1 - S^A}$. It then follows immediately from (48) and (49) 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. ∎

49