On-the-Job Skills Acquisition and Growth Miracles

Paul Beaudry
Department of Economics, UBC

Patrick Francois
Department of Economics, UBC and CentER Tilburg

January 2004

Abstract: The present paper explores on the job skills acquisition as an engine of growth in GDP per capita. It is posited that such skills acquisition occurs as a natural by-product of the production process. The analysis demonstrates that when these are the unique source of productivity growth, such on the job skills, even if fully rewarded, and competitively provided, may be consistent with negligible returns to measured experience. The contribution of such an avenue will thus not generally be correctly measured by standard growth accounting procedures, which will instead suggest productivity differences that arise due to underlying technology differences. It is argued here that on the job skills as a source of development generate predictions that are consistent with the empirical record, and that explains cross country GDP per capita differences without the need to posit a disembodied factor, and without the need to posit implausibly large externalities. It is also consistent with observed migration patterns.

1 Introduction

Few contemporary economists would dispute the central role played by technological change as the engine of long-term growth in GDP per capita. Technology is often viewed as a disembodied set of designs, inventions or processes, which should be relatively simple to transmit across countries. But the relatively clear evidence that this seemingly simple transmission does not occur remains something of a puzzle, suggesting to some the existence of pernicious barriers to technology mobility.[1]

Microeconomic studies of technology acquisition in LDCs however, point to a more complex path of transmission which may be difficult even in the absence of technological barriers. Briggs, Shah and Srivastava (1995, p.7), in a comprehensive study of sub-Saharan African manufacturing emphasize the role of individuals in both the transmission and diffusion of technology

> "A large part of technology involves uncodified knowldege: rules of thumb acquired only with experience and via sustained interaction with the people and institutions embodying this know-how. Both within countries and transnationally, virtually all studies of technology transfer find that the diffusion of technology results, in the main, from movements or interactions of individuals from firm to firm and from country to country. There appears to be *no* evidence that the transfer of technology can be achieved effectively by other modes independently, such as via technology licenses, or for that matter, blueprints sent over the internet unaccompanied by sustained individual interaction." Their italics.

William Easterly (2002) describes one episode of such a technology exchange. In 1979 Daewoo signed a collaborative agreement with Desh Garment Ltd in Bangladesh. Daewoo agreed to bring 130 Bangladeshi workers to Korea for training at a Daewoo plant and in return Desh would pay commissions amounting to 8 percent of sales value. 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,

[1] Explaining this puzzle has motivated the positing of, and search for, "barriers" to technology adoption, see for example Parente and Prescott (2003).

and during the 1980s, of the 130 workers initially trained by Daewoo, 115 of them had left Desh to set up their own garment export firms. Today Bangladesh exports $2 billion in garments.

Clearly a large part of successful technology implementation is played by these "rules of thumb acquired only with experience", a process which we here refer to as on-the-job skills acquisition. A dramatic case of the productivity improvements to be had by the acquisition of such on-the-job skills has already been recounted by Lucas (1993). Drawing on studies by Searle (1945) and Rapping (1965) he documents the productivity rise in the manufacture of the "Liberty Ship" in 14 US shipyards during World War II. Over this four year period, each doubling of cumulative output reduced man-hour requirements, per ship, by between 12 and 24 percent. The increases were even greater on a per-yard basis. Similar documentation of productivity growth through on-the-job knowledge acquisition is cited for a two year period in the famous Swedish Horndal steelworks by Arrow (1962), and in the production of airframes by Alchian (1963).

Doubts that the acquisition of such on-the-job knowledge could underlay productivity increases over sustained periods are dispelled by the account of the Lawrence cotton textile company #2 Mill in Lowell Massachusetts, examined by Paul David (1975). Machines in the mill were installed in 1834 and company records indicated no machinery replacement before 1856. Thus, this example provided a relatively rare case of fixed production with fixed facilities over a longer period. It was found that over the whole period average labor costs exhibited a sustained decline. Compounded labor productivity growth ran at about 2% per annum. Controlling for secular variation due to capacity variation by examining productivity from peak to peak, David recounts initial growth rates of 3.37% from 1839-46, 2.35% from 1846-51, and .29% from 1851-56.

To economists concerned with understanding persistent cross-country differences, such examples are instructive. For some reason, in some places, at some times, production seems to have been organized in a way that enabled people to rapidly transmit accumulated knowledge to one another. Such transmission has facilitated, often dramatic, increases in productivity. We believe this process may play a central, and relatively ignored, role in understanding both development successes and failures. We find the examples of skills acquistion accompanying technological change, and in fact facilitating technological change, a likely ingredient of the growth miracles of the East Asian economies. Annual growth rates of 6-7% in per-capita GDP seem to suggest rapid dissemination of existing knowledge, not its rapid invention. And we believe this catch-up seems plausibly to have necessitated a large amount of labor skilling on-the-job.

Cross-country evidence seems to discount the role of skilling through formal channels, (schools or universities) or the importance of physical capital in explaining cross-country differences; see for example Klenow and Rodriguez-Clare (1997) for a multi-region perspective, and Devrajan, Easterly and Pack (2003) for a focus on Sub-Saharan Africa. It has also been argued that skills acquired on the job is similarly unpromising, especially if returns to these skills are appropriated by labor. On the job skills proxied by experience do not seem to vary sufficiently to be able to explain a large part of cross country variation, see Klenow and Rodriguez-Clare (1997). The present paper aims to demonstrate that this is not the case; we will show that even when on the job skills are the *only* source of productivity differences across country, and even when the returns to such skills fully accrue to labor, the implied pattern of observed returns to experience is entirely consistent with the empirical record.

There are three main problems with the view that embodied human capital acquired on the job, is at the heart of understanding cross country productivity, or income level, differences. These are: 1) Cross-country growth regressions demonstrate returns to experience are tiny. This is true also in time series analysis. 2) Given tiny returns to experience, and the negligible impact of formal education on growth, there must be huge externalities to this sort of on-the-job skills acquired. This is problematic because estimated private returns to education are far smaller than these external effects would need to be; estimated externalities from formal education appear negligible. So these externalities to on-the-job skills must far outstrip any other sort of knowledge benefit known to be created. 3) If this sort of experience related skill is at the heart of cross country level differences, incentives for international migration so created would appear to be greatly at odds with actual migration patterns.[2] In particular, the skilled in rich countries should have great incentives to move to the poor countries, and the skilled in the poor should stay put, and certainly should not wish to migrate to unskilled positions in rich countries. Moreover, there should be huge skill premia in poor countries relative to the rich.

In contrast, we will argue: 1) Returns to experience can be arbitrarily small, and even zero, even if on-the-job skills learned through experience are the source of cross-country productivity differences, and even if the accumulation of such skills underlay a particular country's growth experience; 2) This can occur even if labor is fully able to appropriate all of the benefits generated from its own productivity, i.e., there need be no externalities; 3) On-the-job skills acquisition is

[2]This point has been made in numerous places, Romer (1995) is one example.

entirely consistent with observed migration patterns. In particular, it is consistent with everyone, both skilled and unskilled, preferring to locate in rich countries.

Though of undeniable importance, we will not examine the microeconomic or institutional features that have allowed some places at some times to be able to undertake this sort of rapid knowledge dissemination. Our analysis is instead limited entirely to the macroeconomic implications of such a process of on-the-job skills acquisition. Accordingly, for our purposes, economies will differ only in at least one of three fundamentals: the rate at which their members discount the future, the rate at which embodied skills decay (a factor strongly related to health, mortality and stability of production) and the productivity of the underlying technology freely available to them. For much of our analysis, this third factor will be assumed equivalent across countries, that is there will be assumed to be zero barriers to technology adoption.

We will assume throughout that all economies have the institutional make-up required to allow on-the-job skills to be disseminated. In fact, we will assume that competitive factor markets ensure this is done efficiently. Workers will be fully rewarded for their productive contribution to output, and workers will be fully rewarded for the impact their skills have on the future skills of their co-workers. Moreover, workers will be assumed to be unconstrained in their capacity to finance the acquisition of such skills.

Our model's externality and market failure free environment is not meant to be a realistic depiction of the world. In fact, a conclusion of our analysis is that the way economies manage on-the-job skill transmission is important, and deserves careful scrutiny. In our framework this skilling is THE source of growth (as technology movement is barrier free), and is THE key to development. We seek only to show that assuming a process where this factor is the key, has implications consistent with the empirical record. We believe it may well be a crucial factor that has been ignored, but only further study will determine this.

Relation to Previous Studies

Lucas (1993) also highlighted the role of on-the-job learning in his seminal paper on growth miracles. He interpreted this on-the-job learning as indicating an externality to human capital, and, along the lines of Stokey (1988) and Young (1991), built a model in which skills acquisition, for which he used the term "learning-by-doing", raised the proficiency of labor and allowed expansion of variety. The rate at which new goods could be incorporated into production was linked to the level of proficiency, or learning, in existing varieties. The externality arose because increased variety raised the productivity of all labor. These skills thus provided an explanation for why technology (in the form of new goods) could not readily move across countries. As such, it could be interpreted as providing a micro foundation for the disembodied factor seen to be important from the growth accounting.

Our focus is not on the impact of skills acquisition on advances in technology as such. A more straightforward channel is studied here: put simply, when a new technology is introduced, productivity rises because workers, by working with others who know how to use it, become more skilled in its use. On-the-job skills acquisition thus raises productivity directly. The type of externality posited by Lucas (1993), i.e. increasing worker skills learned on the job facilitates introduction of more productive technology, is not allowed here, and is actually not evidenced in his "Liberty Ship" example, nor in the case studies exhibiting on-the-job skills that we cite. The simple reason being that in order to isolate the impact of skills improvement on output, these studies necessarily analyze environments in which technology is stationary. Although it seems emimently plausible that the channel Lucas studied is important, the evidence provided actually supports the much more simple conduit, which we examine in the present paper.[3]

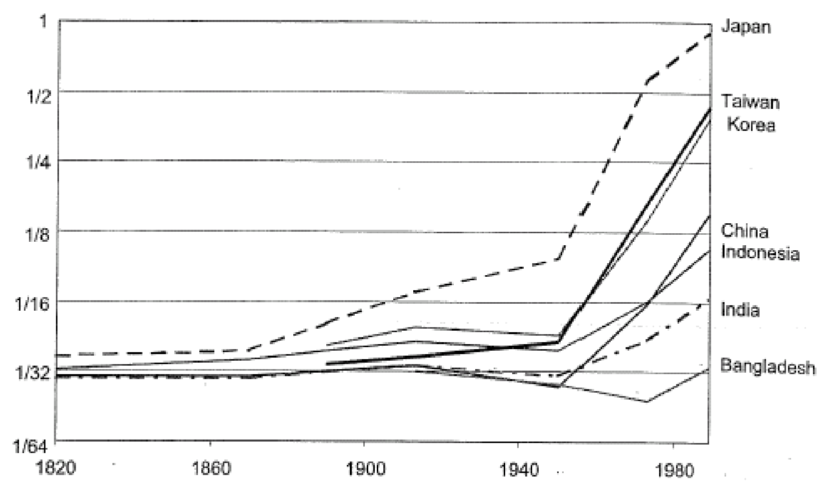
An important implication of our treatment is that, since learning directly raises the productivity of workers, its benefit is embodied in them. Even if workers cannot contract over their future earnings, there is nothing, in principle, which necessitates this being an externality. As Becker's (1975) theory of human capital acquisition made clear, if learning yields benefits embodied in workers that are transportable to other firms, in a competitive economy, workers rather than firms will pay for these. If workers become skilled by working with those already skilled, then training is excludable and chargeable, and can be simply analyzed in a decentralized competitive equilibrium.

Our emphasis on skills acquired through use of a new technology bears much in common with 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 expanded each period, they demonstrated the existence of a stationary distribution of technologies in use (with constantly increasing average productivity) and constantly increasing wages. Of closer relevance for the present study, they found, as we do here, a flattening of the wage-earnings profile

[3] The "worker as knowledge conduit" view has received empirical support in the US by Glaeser et. al (1992).

as a technology aged. 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 an old technology to a newer one of higher productivity. As our application is explicitly to developing countries, and the problem of persistently lower living standards, we determine the conditions under which economies will not utilize the superior technology, and the development path they will follow in its use if, in contrast, they do develop it.

Our paper is thus focused differently than theirs. We believe that the growth miracles are best understood as once off phenomena that can occur when an economy undergoes large scale modernization. As the cross country evidence shows, any aggregate convergence in world income that has occurred over the last century has not been due to persistent growth rate advantages of poor countries over rich. On the whole, patterns of growth rates in GDP per capita between poor and rich countries have not varied significantly since the second world war. Instead, the aggregate convergence observed over the last 50 years is driven almost entirely by a handful of countries that have managed to rapidly (in the space of one generation) revolutionize their economies. Parente and Prescott's (2002) Figure 2.2 reproduced below, plots the income per capita paths of countries (relative to the 1985 US level) that have experienced miracles, using data from Maddison (1995).



Parente and Prescott's (2002) Figure 2.2

It is these rapid episodes of revolutionary change in GDP per capita which the possibility of technological catch-up affords, that are the focus here. Consequently, rather than a constantly increasing set of technologies in steady state, as in Chari and Hopenhayn, we focus on modeling the transition that occurs when a country moves into modern production. That is, we aim to understand the potential process of upgrading to a wholly new mode of production that is represented by these dramatic transition paths. We believe that this is a more correct analog for the potential underlying the rapid growth experiences that have been observed in the empirical record. Our principal aim is to derive the implications of this process for measured returns to these skills, in order to determine whether such a process of on the job skills acquisition can be consistent with the empirical record.

The emerging consensus view of the growth accounting literature is that differences in observable factors - physical and human capital - cannot explain level differences in income per capita. A not directly observed factor, thought to be disembodied, is playing a key role. The need to treat the claimed factor of importance as disembodied leads to problems of interpretation for the endogenous growth literature explaining development differences and, particularly explaing why most LDCs have not been able to rapidly converge to levels in the West. Theories that emphasize a non-rivalrous technology (such as Arrow (1962a), Shell (1966), Grossman and Helpman (1991) and Aghion and Howitt(1992)) have had to explain why the technology seems so hard to implement in poor countries. One explanation has been "barriers to adoption", such as legal, institutional, socio-cultural or political factors, see Parente and Prescott (2002). No doubt, such factors are important, and probably critical in some instants. But with upto thirty-fold level differences in GDP per capita between the rich and poor, these barriers to adopting from the West (even greatly outdated technologies) would have to be huge in order to explain the cross country variation.[4]

On the other hand, theories emphasizing human capital (for example, Arrow (1962b), Lucas (1988) and Barro and Sala-i-Martin (Ch. 5)) depend critically on externalities generated by human capital, since observed returns to human capital are not large enough to explain cross-country experiences. These theories need to account for why externalities disperse freely within countries, but do not seem able to cross borders. Externalities embodied in the human capital

[4] Parente and Prescott (2002) argue that a significant barrier is labor and entrenched interests' resistance to the reorganization that necessarily accompanies technological upgrading. Though it is not possible to obtain cross country estimates of this, they compute the implied level of such resistance required to explain the cross country distribution. The implied resistance they find would have to be of an order of magnitude which is equivalent to 42% of an industry's annual value added, to explain this variation.

stock, but not the individual, would work, but must be large to explain the observed cross-country distribution. Klenow and Rodriguez-Clare (1997) estimate that social returns to an extra year of human capital need to be on the order of 20%, a full 5% higher than even the most optimistic estimates of private benefits in order to explain these cross-country differences. However, estimates of the social impact of human capital, suggest these are negligible.

The paper proceeds as follows. Section 2 sets up a model examining on the job skills acquisition in a competitive economy. The steady state and transitional dynamics of such a process are explored, and it is shown that the on-the-job skills framework can account for arbitrarily large differences in income per capita across countries. Section 3 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 growth accounting results are completely consistent with observed returns to experience being arbitarily small. Section 4 then explores a slight extension of the model, which can be used to address migration patterns. The framework will be able to explain why unskilled labor that moves from poor to rich countries will reap high returns, whereas similar incentives may not exist for the movement of skilled labor to poor countries. Section 5 concludes.

2 The Model

Households

Consider an economy comprising overlapping generations of individuals which experiences zero population growth. Each individual is risk neutral and inelastically supplies one unit of labor. Let $\delta > 0$ denote the per instant probability of an individual leaving the model economy (death or exogenous transition) and $\rho > 0$ be the instantaneous discount rate. The labor force is of constant size 1, so that a new cohort of workers enters each instant of size δ, i.e. $\int_{-\infty}^{t} \delta e^{-\delta(t-s)} ds = 1$. All newly entering workers are unskilled, where skills will be defined below. Discounted expected lifetime utility is

$$U = \int_0^\infty c\,(t)\, e^{-(\delta+\rho)t} dt,$$

where $c\,(t)$ denotes consumption.

Production

In the first section we will model production as comprising two types of activity; denoted traditional and modern. Modern production occurs under constant returns to scale and utilizes: S, which denotes the proportion skilled in use of the modern technology and U, which denotes the proportion without skill working with the modern technology. The term L denotes the proportion working in the traditional technology, which also exhibits constant returns to scale in this single input. We wish also to allow for scaling of the modern technology by a (potentially country-specific) Harrod-neutral efficiency parameter, θ. We may think of this as reflecting a country specific implicit tax on the modern sector or, more directly, as reflecting the relative efficiency of the new technology over the old.

Total output produced in this economy is:

$$\begin{aligned} Y &= \theta F(S,U) + H(L) \\ &\quad \text{and since } H \text{ is linearly homogeneous:} \\ Y &= \theta F(S,U) + \overline{w}L, \end{aligned}$$

where $\overline{w}$ is labor's product in the traditional technology. An alternative interpretation of the traditional technology is as one in which skills are already abundant in the population, so that the S, U distinction disappears (as discussed further below). Note also that capital plays no role in the analysis, this is without loss of generality, as we discuss further subsequently. We impose a type of Inada condition on production:

$$\begin{aligned} \text{for } U &> 0, \ \lim_{S\to 0} F_1(S,U) = \infty \text{ and } F_2(0,U) = 0 \\ \text{for } S &> 0, \ \lim_{U\to 0} F_2(S,U) = \infty \text{ and } F_1(S,0) = 0. \qquad (1) \\ \text{by continuity then } \exists\, S' &: \ F_1\left(S', 1-S', \right) = F_2\left(S', 1-S'\right). \qquad (2) \end{aligned}$$

In the Cobb-Douglas case, this is satisfied with $S' = \alpha^{\alpha}(1-\alpha)^{1-\alpha}$. A natural interpretation of this condition is that production necessitates some individuals being placed in positions requiring skills, but when all such positions are occupied, extra skilled workers are more usefully allocated to positions that could as easily be filled by the unskilled. When this occurs, the productivity differential between a skilled and unskilled worker disappears. Under this interpretation, the traditional technology has already reached its level of S', so that the distinction between skilled and unskilled in this technology is immaterial. Clearly, for it to ever be the case that the traditional technology may give way to the modern, we require that if the modern technology is not skill constrained, the unskilled are more productive working with it than in traditional:

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

On-the-job skills acquisition

If an individual works in the new technology, he becomes skilled randomly following a Poisson process with constant instantaneous rate Ω. We will explore a slight generalization of this acquisition process in Section 4.

As argued earlier, though on-the-job skills acquisition is a by-product of production, there need be no externalities in this process. This is because skills acquisition is excludable; individuals who do not work do not acquire skills. Thus we assume simply that factors are rewarded competitively at their marginal products, as in a standard Walrasian equilibrium. Denote skilled wages at time t, w_t^s and for the unskilled, w_t^u. Skilled workers are fully rewarded for their role in disseminating skills to the unskilled in this set-up through equilibrium entry. To see why, suppose that skills acquired on the job are extremely valuable. Then, unskilled workers are willing to accept low wages to (probabilistically) become skilled in the new technology. This leads to high entry of unskilled workers into new production, lowering their marginal product, and hence the unskilled wage. The impact on the skilled wage is the converse. An instantaneous equilibrium is achieved when the net present value of the expected returns to learning, plus the instantaneous wage, just equal a worker's opportunity wage. This is a forward looking condition which anticipates the economy's future path of skills acquisition and hence the returns to being skilled. At this point, we do not consider the institutional structure that allows the pricing of such on-the-job skills but will return to this issue in the concluding discussion.

Since we analyze a competitive equilibrium, the solution to the planner's problem is equivalent to the outcome from a decentralized economy, so that we consider the former in what follows.

The Planner's Problem

The planner's problem amounts to allocating skilled and unskilled workers between production in both the traditional and modern technology at each instant. The planner takes the initial stock of skilled as given, and in most formulations we shall assume this is zero, but it then evolves according to the skills acquistion technology. The skilled can work in one of three tasks: they can remain in the old technology receiving $\overline{w}$, denote this proportion $\overline{\gamma}$, they can work as skilled in the new technology, denote this proportion γ, or they can work as unskilled in the new technology, the remaining $(1 - \gamma - \overline{\gamma})$. The unskilled have fewer options. They can work as unskilled in the new technology where they probabilistically acquire skills at rate Ω, proportion β, or they can work in the old technology and remain unskilled for sure, the remaining $(1-\beta)$.

It turns out to be simpler to analyze the optimization problem from the perspective of a single cohort. Accordingly denote the levels of $\gamma, \overline{\gamma}$ and β chosen at time t by a cohort born at time $q \leq t$ as $\gamma(q,t)$, $\overline{\gamma}(q,t)$ and $\beta(q,t)$. Denote the cohort's remaining size at t as $N(q,t)$ and its measure of skilled $S(q,t)$.

The cohort pools income so that each member has equal consumption.[5] Its optimization problem is:

$$\begin{aligned}
& \max_{\gamma,\overline{\gamma},\beta} \int_q^\infty \left(\frac{C(q,t)}{N(q,t)}\right) e^{-(\delta+\rho)(\tau-q)} d\tau \\
& s.t. \\
C(q,t) &= w_t^s \gamma(q,t) S(q,t) + w_t^u \left[\begin{array}{c} (N(q,t)-S(q,t))\beta(q,t) \\ +(1-\gamma(q,t)-\overline{\gamma}(q,t)) S(q,t) \end{array}\right] \\
& \quad + \overline{w}\left[(1-\beta(q,t))(N(q,t)-S(q,t)) + \overline{\gamma}(q,t) S(q,t)\right] \qquad (4) \\
\dot{S}(q,t) &= (N(q,t)-S(q,t))\beta(q,t)\Omega - \delta S(q,t) \qquad (5) \\
N(q,t) &= e^{-\delta(t-q)} \qquad (6) \\
\{\gamma(q,t), \overline{\gamma}(q,t), \beta(q,t)\} &\in [0,1] \ \forall t. \qquad (7)
\end{aligned}$$

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:

$$\begin{aligned}
\frac{\dot{\chi}(q,t)}{\chi(q,t)} &= \left(\frac{1}{\chi(q,t)} - 1\right)\beta(q,t)\Omega \\
\Rightarrow \dot{\chi}(q,t) &= (1-\chi(q,t))\beta(q,t)\Omega \qquad (8) \\
& \quad and \\
c(q,t) &= w_t^s \gamma(q,t)\chi(q,t) + w_t^u \left[(1-\chi(q,t))\beta(q,t) + (1-\gamma(q,t)-\overline{\gamma}(q,t))\chi(q,t)\right] \\
& \quad + \overline{w}\left[(1-\beta(q,t))(1-\chi(q,t)) + \overline{\gamma}(q,t)\chi(q,t)\right]. \qquad (9)
\end{aligned}$$

The cohort's problem can now be re-written as:

$$\begin{aligned}
& \max_{\beta,\gamma,\overline{\gamma}} \int_0^\infty c(q,t) e^{-(\delta+\rho)(\tau-q)} d\tau \\
& \text{s.t.} \quad (6)\,(7)\,(8) \text{ and } (9).
\end{aligned}$$

[5] Since individuals are risk-neutral, the solution to the planner's problem for the pooled cohort is equivalent to that for each individual.

The first order conditions of this problem imply:

$$\overline{w} - w_t^u \begin{cases} > \lambda_t \Omega & \text{if } \beta_t = 0 \\ = \lambda_t \Omega & \text{if } \beta_t \in (0,1) \\ < \lambda_t \Omega & \text{if } \beta_t = 1 \end{cases} \tag{10}$$

$$w_t^s - w_t^u \begin{cases} < 0 & \text{if } \gamma = 0 \\ = 0 & \text{if } \gamma \in (0,1) \\ > 0 & \text{if } \gamma = 1 \end{cases} \tag{11}$$

$$\overline{w} - w_t^u \begin{cases} < 0 & \text{if } \overline{\gamma} = 0 \\ = 0 & \text{if } \overline{\gamma} \in (0,1) \\ > 0 & \text{if } \overline{\gamma} = 1 \end{cases} \tag{12}$$

$$\dot{\lambda}_t - (\delta + \rho)\lambda_t = \beta_t w_t^u + (1 - \beta_t)\overline{w} - w_t^s + \lambda_t \beta_t \Omega \tag{13}$$

where λ denotes the co-state variable, and where the w^i are dependent upon S, which we return to subsequently.

Note firstly that the optimal values of β, γ and $\overline{\gamma}$ are pinned down entirely by the relative value of wages, and are thus independent of the cohort's state variable, χ. Since this implies the optimal β will be equivalent across cohorts, we have used the less cumbersome notation, β_t, above and from hereon, with subscript t denoting time. This also implies that wages can be expressed directly in terms of aggregates. So we denote $w_t^s = \theta F_1(\gamma_t S_t, \beta_t (1 - S_t))$ and $w_t^u = \theta F_2(\gamma_t S_t, \beta_t (1 - S_t))$.

Since w_t^s and w_t^u are market clearing wages determined by the evolution of aggregate skills, not skills within the cohorts, we must determine the variation in the aggregate stock of skilled workers, S_t. Integrating across cohorts yields:

$$\begin{aligned} S_t &= \int_0^t \chi(q,t) L(q,t)\, dq \\ \Rightarrow S_t &= \int_0^t \chi(q,t) e^{-\delta(t-q)} dq \\ & \quad \textit{and} \\ \dot{S}_t &= (1 - S_t)\beta_t \Omega - \delta S_t. \end{aligned} \tag{14}$$

The system is thus described by the first order conditions determining β and γ, equations (10) and (11) and the two differential equations (13) and (14).

2.1 Analysis

Figure 1 depicts the λ, S space being cut by two functions that are critical for the evolution of the system. The first is the vertical line S', defined by the equality in expression (3). To the right of this line $\gamma < 1$ from (2), and to its left $\gamma = 1$.

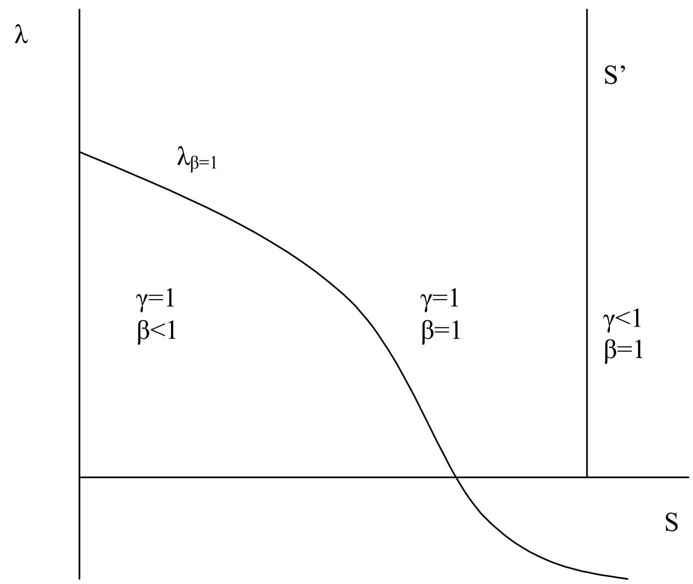


Figure 1

This follows from (11) and the following lemma.

Lemma 1: *Skilled workers never use the traditional technology;* $\overline{\gamma} = 0$ *for all* t.

Intuitively, since the modern technology is superior by construction, returns to working in it, even as an unskilled worker, must dominate remaining in the traditional. The second important function is that sketching $\lambda = \frac{\overline{w} - \theta F_2(S,(1-S))}{\Omega}$, derived from condition (10), denoted as $\lambda_{\beta=1}$ from now on, which traces the points at which $\beta = 1$. Below this function $\beta < 1$.

2.1.1 Dynamics in $\beta = 1$ region

For $\beta = 1$ we have:

$$\begin{aligned} \dot{\lambda} &= (\delta + \Omega + \rho)\lambda + \theta \left[F_2 (S, 1-S) - F_1 (S, 1-S) \right] \\ \dot{S} &= \Omega (1-S) - \delta S. \end{aligned} \tag{15}$$

In this region, the $\dot{\lambda} = 0$ function is a downward sloping line asymptoting at $S = 0$ and meeting the S axis at $S = S'$, as depicted in Figure 2. Above the S' locus, $\gamma < 1$, and λ must take value 0. For $\beta = 1$, the $\dot{S} = 0$ locus is the vertical line given by $S = \frac{\Omega}{\Omega + \delta}$.

It is possible that the $\dot{\lambda} = 0$ function intersects the $\lambda_{\beta=1}$ function. This occurs at values of S defined by.

$$\frac{\theta \left[F_1 (S, 1-S) - F_2 (S, 1-S) \right]}{\delta + \Omega + \rho} = \frac{\overline{w} - \theta F_2 (S, 1-S)}{\Omega}. \tag{16}$$

It turns out that an implication of the following lemma is that these functions necessarily intersect at most twice.

Lemma 2: $S = \frac{\Omega}{\Omega+\delta+\rho}$ *uniquely solves* $\min_S \frac{\theta(F_1(\cdot)-F_2(\cdot))}{\delta+\Omega+\rho} - \frac{\overline{w}-\theta F_2}{\Omega}$.

Generically, either such points of intersection do not exist, or there exist two such points of intersection. (We ignore the measure zero parameter configurations yielding a tangency). When the $\dot{\lambda} = 0$ locus does not cross the $\lambda_{\beta=1}$ in the positive orthant, the dynamics are always governed by the $\beta = 1$ equations (15). This case is depicted in Figure 2 below, and, as the directional arrows indicate, movement in the system is straightforward.

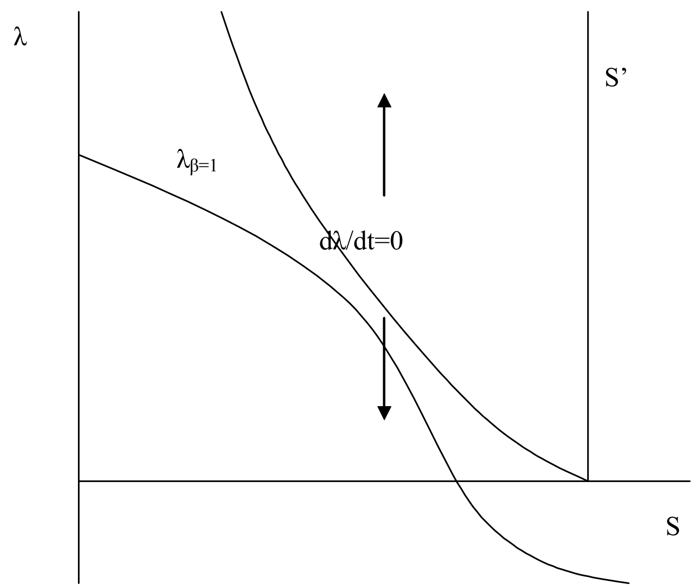


Figure 2

When crossing points exist, denote them by S_1 and S_2 with $S_1 < \frac{\Omega}{\Omega+\delta+\rho} < S_2$. Thus, along the $\dot{\lambda} = 0$ function, for $S \in (S_1, S_2)$, $\beta < 1$. This is analyzed in the next section.

It is also possible for there to be a unique intersection between $\dot{S} = 0$ in (15) and $\lambda_{\beta=1}$, which is defined by:

$$S = \frac{\Omega}{\Omega+\delta}, \lambda = \frac{\overline{w} - \theta F_2\left(\frac{\Omega}{\Omega+\delta}, \frac{\delta}{\Omega+\delta}\right)}{\Omega}.$$

When a crossing does not exist in the positive orthant, i.e. when $\overline{w} - \theta F_2\left(\frac{\Omega}{\Omega+\delta}, \frac{\delta}{\Omega+\delta}\right) < 0$,the movement of the system is always towards $S = \frac{\Omega}{\Omega+\delta}$ as depicted in Figure 3 below. Conversely, a crossing exists and for $\lambda < \frac{\overline{w}-\theta F_2\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)}{\Omega}$, the $\dot{S}=0$ relationship implies $\beta<1$, which we analyze in the next section.

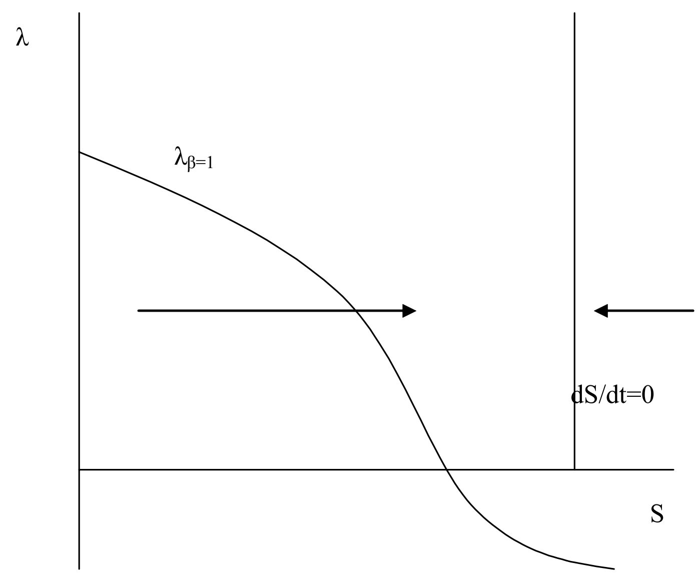


Figure 3

2.1.2 Dynamics in the $\beta < 1$ region

For $\beta < 1$ we have:

$$\begin{aligned} \dot{\lambda} &= (\rho+\delta)\,\lambda + \overline{w} - \theta F_1\left(S, \beta\left(1-S\right)\right) \\ \dot{S} &= \Omega\left(\cdot\right)\left(1-S\right)\beta - \delta S \end{aligned} \tag{17}$$

The $\dot{\lambda}=0$ locus is defined for $\beta<1$ only if the points S_1, S_2 exist. It is defined by:

$$\overline{w} = \theta F_1\left(S, \beta\left(1-S\right)\right) - \left(\delta+\rho\right)\lambda. \tag{18}$$

It follows from homotheticity that such locuses are horizontal lines.

Lemma 3: *For* $\beta<1$, $\dot{\lambda}=0$ *implies* $\frac{d\lambda}{dS}=0$.

Given this lemma, the existence of two intersection points, S_1 and S_2 implies that, beneath the $\lambda_{\beta=1}$ locus, two horizontal lines define $\dot{\lambda}=0$ from (18). This is the case depicted in Figure 4.

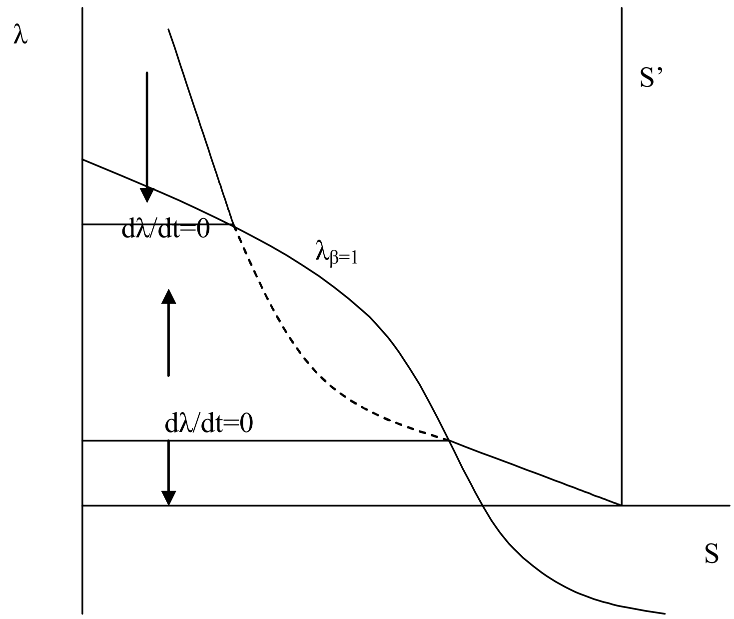


Figure 4

Turning finally to the the $\dot{S} = 0$ locus for values beneath the $\lambda_{\beta=1}$ line. The following lemma shows that it too is a horizontal schedule.

Lemma 4: *For* $\beta < 1$, $\dot{S} = 0$ *implies* $\frac{d\lambda}{dS} = 0$.

Figure 5 depicts the phase space when such a point of intersection exists.

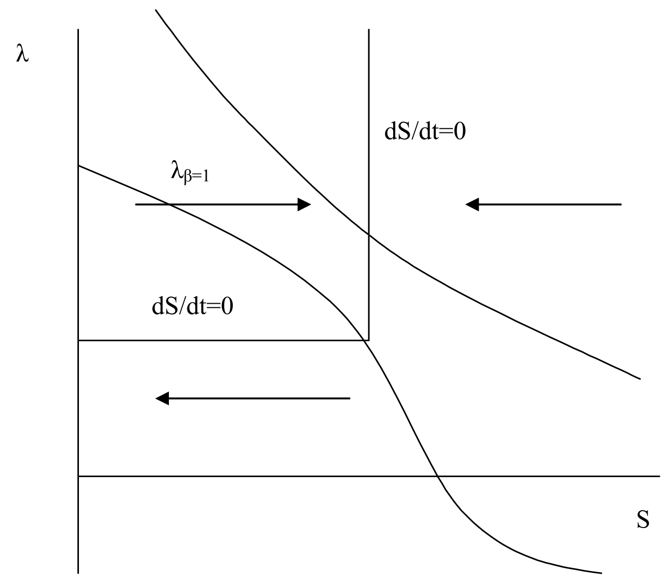


Figure 5

Combining these possibilities implies four mutually exclusive outcomes can occur: outcome A) neither locus crosses the $\lambda_{\beta=1}$ line. This is depicted in Figure 6; outcome B) both the $\dot{\lambda} = 0$ and the $\dot{S} = 0$ loci cut the $\lambda_{\beta=1}$ line, depicted in Figure 7; outcome C) Only the $\dot{S} = 0$ locus cuts the $\lambda_{\beta=1}$ line, depicted in Figure 8 and outcome D) only the $\dot{\lambda} = 0$ locus cuts the $\lambda_{\beta=1}$ line, or if the $\dot{S} = 0$ line also cuts, it cuts at $S > S_2$. This is depicted in Figure 9 for the case of the two different possibilities of $\dot{S} = 0$.

2.2 Steady States and Transitional Dynamics

We analyze each configuration represented in Figures 6-9 seperately. Figure 6 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, but a skill premium will persist. A qualitatively similar situation occurs when the $\dot{S} = 0$ locus is located to the right of the $S = S'$ locus. This case, depicted by the dotted line for $\dot{S} = 0$ in the figure, identically has a unique steady state, but in it there is no skill premium; $\lambda = 0$. In this case steady state skills in the modern technology are accumulated to the point of abundance.

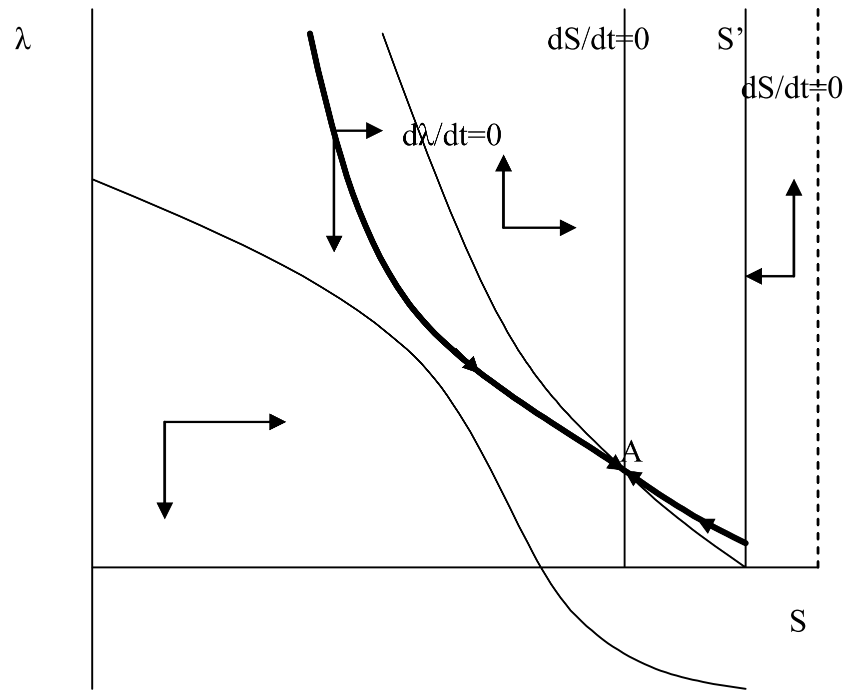


Figure 6

Figure 7 depicts the phase space for the situation corresponding to crossing of both loci with the $\lambda_{\beta=1}$ line. Here again, the steady state reached is independent of the economy's starting position. Now, however, the economy remains fully mired in the traditional technology. In steady state, though there is a positive skill premium ($\lambda > 0$) 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. In such an economy, skill premia in the modern technology will be steadily rising as it is approaching the steady state. The economy is unable to sustain modern production because δ is too high and Ω too low to maintain a stock of skilled individuals from whom the unskilled can learn. Any such stock that is exogenously created, say through an explicit government training program, or through migration from abroad, will eventually be depleted without triggering large scale development of the modern technology.

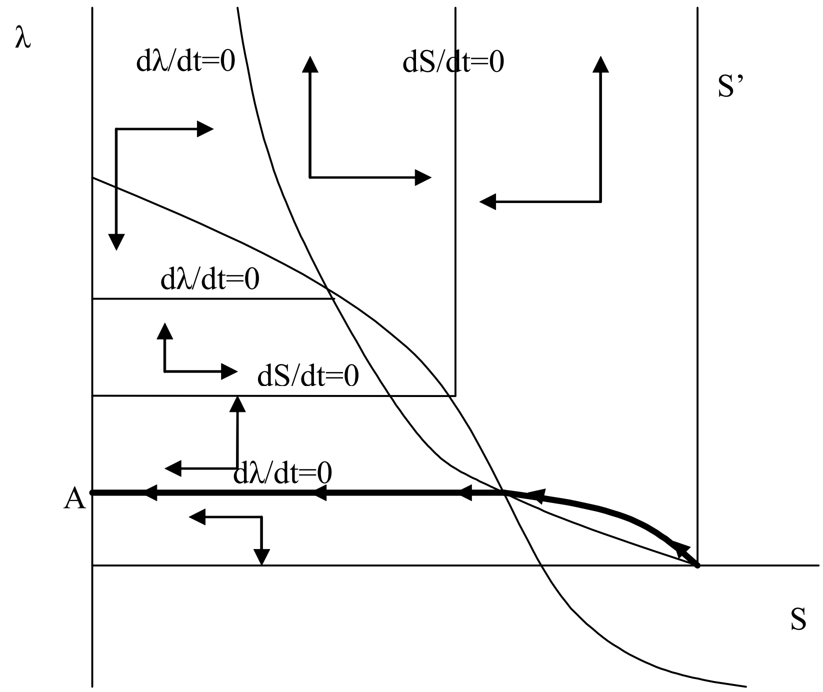


Figure 7

Figure 8 depicts a situation in which only the $\dot{S} = 0$ locus intersects $\lambda_{\beta=1}$. Here again the unique steady state of the system is at the point A where all individuals work in the modern technology, and there remains a positive skill premium. Transitional dynamics are similar to those in Figure 6. Even an economy inheriting zero skills will eventually converge on a steady state with fully modernized production. This is also an economy with relatively high δ and low Ω, as indicated by the location of the $\frac{dS}{dt} = 0$ locus, however the distinction from the situation in Figure 7 would arise from the economy below having a lower value of ρ. If low enough, as depicted below, the $\frac{d\lambda}{dt} = 0$ locus shifts out enough to ensure an intersection of the $\frac{dS}{dt} = 0$ function in the region $\beta = 1$. Intuitively, the future is valued highly enough to offset the relatively low possibility of learning the new technology. This ensures a steady state where all unskilled workers are profitably placed into the new technology, and a sufficiently large stock of skilled is maintained.

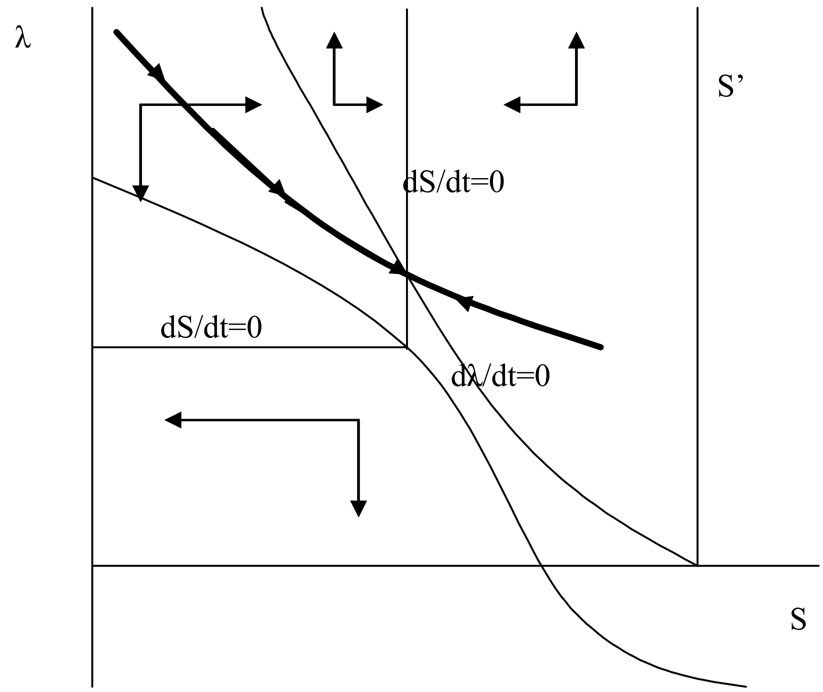


Figure 8

Finally, Figure 9 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 $S > S_2$. 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, but the introduction of even an arbitrarily small number of skilled would lead to production with the modern technology. With the on the job skills accumulated through that production, the numbers skilled would increase, and so too would modern production, until eventually all production is modern, and the point A is reached. Note also that in this figure there is a qualitatively different case depicted for the dashed $\dot{S} = 0$ line denoted $\dot{S} = 0(2)$. This occurs when the $\dot{S} = 0$ locus does not intersect the $\lambda_{\beta=1}$ line. 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.

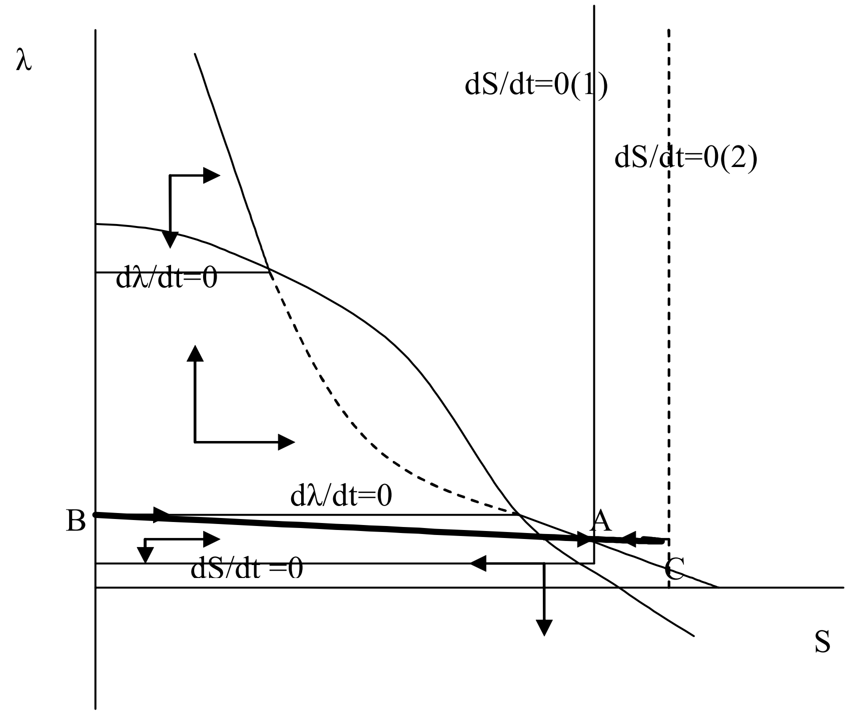


Figure 9

In situations depicted in Figures 6 and 8 the economy converges to steady states which may or may not have skill premia but in which the traditional technology is completely shut down. Skills may or may not be in full abundance. For the situation in Figure 7, conversely, the traditional technology remains in full use and no skills are accumulated, even though there is a positive premium associated with being skilled. Recall that this outcome (persistence of traditional production) occurs only if the $\dot{\lambda} = 0$ locus intersects the $\lambda_{\beta=1}$ locus. Finally, the situation in Figure 9 implies hysteresis. Economies without skill will not develop the modern technology, but even small amounts of skill will eventually lead to full diffusion. Such an economy would experience rapid growth in light of even a small initiative that was able to introduce some skilled workers.

The ensuing situation for any particular economy depends critically on the relative locations of the loci sketched above. Since these are determined by the values of exogenous parameters, θ, δ and ρ, we have sharp predictions regarding the technology that will be used in steady state, and for initial levels of skills, the economy's transition paths:

Proposition 1: *For a given productivity level in the modern technology θ there exists:*

(a) a critical value, denoted k^*, such that if $\rho+\delta<k^*$, all production eventually occurs using the modern technology.

(b) *a critical value of* δ, denoted δ^*, *such that, if and only if $\delta<\delta^*$ there is no premium to being skilled in steady state; $\lambda=0$.*

(c) *another critical value of δ, denoted δ^{**} such that for $\delta<\delta^{**}$ and any value of ρ, the economy either converges to the fully modern steady state uniquely (that is, when $\delta+\rho<k^*$) or the economy is in the hysteresis case (when $\delta+\rho\geq k^*$). For each value of $\delta\geq\delta^{**}$, there exists a corresponding value of ρ, denoted ρ^* such that, if $\rho>\rho^*$ the economy remains mired in the traditional technology.*

Figure 10 summarizes this proposition's implications for possible steady state outcomes. In the lower left hand corner, low values of δ and ρ, even an economy inheriting zero skills will fully transform to modern production, and because δ is low, there will be no skill premium. The disappearance of a skill premium follows directly from the low labor turnover which ensures steady state skills are persistently high. For continuing low values of ρ and higher values of δ, moving rightward in the graph, the modern technology will still eventually predominate, but the higher values of δ will allow the persistence of a skill premium in the modern steady state. For high values of both δ and ρ, development of the modern technology is not worthwhile. An economy inheriting some skilled workers, or receiving in-migration of them, would utilize the technology, and would also skill some further workers in its use. However, the low rate of skill persistence due to the high labor turnover would eventually see such skills vanish from the population as the economy reverts back to full use of traditional technology. The upper left hand region, high ρ and low δ is a case of hysteresis. The economy's low valuation of the future, high ρ, ensures that, without skills originally present, none would find it worthwhile to incur currently low productivity in order to accumulate them. However, with even a small amount present, they will be used in production, and since skills are persistent, δ is low, this will eventually lead to their diffusion through the population. The final outcome is a full transition to the modern technology.

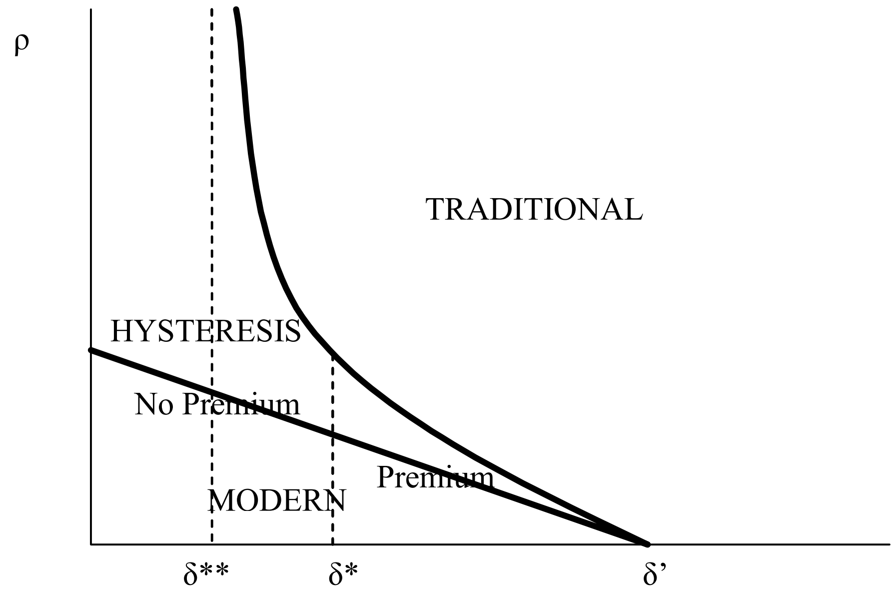


Figure 10

A growth miracle can only occur in transition, and it happens because of catch-up, through the process of on the job skills acquisition. Countries that have already had a growth miracle, or are not going to have one, will not experience growth, unless it is exogenously introduced here. Moreover, countries that will not be able to experience a growth miracle are those with high mortality rates (δ high) low valuations of the future (ρ high) or high implicit or explicit taxes on the modern sector (θ low).

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, in Figures 6 and 8, even an economy without any skills will devote resources to acquiring the new technology. 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 figures is a high valuation of the future. In extreme contrast, Figure 7 depicts an economy with both a relatively low valuation of the future, and a low capacity to both upgrade and maintain skilled labor. 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. A middling case is given in Figure 9. Such an economy will not converge to modern production without some exogenous influx of skilled workers. 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 Implications for Growth Accounting

The implications of this process for growth accounting are now explored. Since skill here is picked up on the job, it will correspond with returns to experience. Such returns are usually estimated from micro data, but how these impact upon growth accounting depends upon whether a cross-section of countries or a time-series is considered:

3.1 Returns to experience in a cross-section

Growth accounting in a cross-section typically involves computing:

$$\ln \frac{Y_i}{L_i} = \alpha_i \ln \frac{EXP_i}{L_i} + \gamma_i \ln \frac{X_i}{L_i} + \ln A_i,$$

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 is usually estimated from a Mincer type earnings regression from 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:

$$\frac{\ln A_i - \ln A_j}{\ln \frac{Y_i}{L}_i - \ln \frac{Y_j}{L_j}}, \tag{19}$$

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 experience is reflected in the returns to the experienced. In order to see how this may be misleading, now suppose that countries i and j have identical values of δ and, the growth process is driven by learning-by-doing, as we have modeled it. For now, assume that $\delta \leq \delta^*$. Assume also that country i has a lower value of ρ so that it lies under the k^* line in Figure 10, whereas country $j's$ value of ρ is so high that growth does not occur - it is above the k^* line.

Now consider an accounting exercise aiming to explain level differences between i and j, using steady state data. Income per-capita is higher in country i, and because its growth is driven by skills, accumulated experience from use of the new technology is entirely responsible for this. However, returns to experience, estimated from the cross-section, are zero. As discussed in Section 2.2, this is because skills are abundant in such a steady state. Consequently, all of the difference in income levels that was caused by differences in experience as a consequence of on-the-job learning is picked up in the term (19) . 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, since it arises in 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 corresponding to $\delta > \delta^*$. 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.

3.2 Returns to experience in a time series

Now consider a single country that is experiencing growth in output per capita according to a process of learning by doing as modeled above. In attempting to pick up the contribution of various factors to the country's growth over time, one would attribute it as follows:

$$\frac{\dot{y}}{y} = \alpha \frac{\dot{\exp}}{\exp} + \gamma \frac{\dot{x}}{x} + \frac{\dot{A}}{A},$$

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

However, experience is 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. But the age and work-force structure is unchanged through the transition. Thus proxying experience by the age structure or time in the labor force indicates that $\frac{\dot{\exp}}{\exp} = 0$. The outcome is thus the same as in the cross-section. Time-series growth accounting exercises will attribute all of the learning by doing to the residual term,

$\frac{\dot{A}}{A}$. This again leads to the erroneous attribution of country $i's$ income growth to an unexplained residual factor instead of its true source; skills purposefully accumulated on the job.

4 Migration incentives

Now consider migration incentives between two countries with access to the same technology but differing in fundamentals and thus located at different points in Figure 10. Suppose that one country, denoted i, has a low value for both fundamentals and in particular that $\delta^i + \rho^i < k^*$, with $\delta < \delta^*$. Whereas for country j, $\rho^j + \delta^j > k^*$ and $\rho^j > \rho^*$. Country i will eventually be in a steady state similar to that depicted in Figure 6, with full use of the modern technology and relatively high productivity and GDP per capita relative to country country j which only uses the traditional technology. Clearly, the unskilled in country j would like to move to country i. There are no skilled in country j in steady state, but the skilled in country i are abundant and thus reap no wage premium over the unskilled in the modern technology, $\lambda^i = 0$. But in the poor country $\lambda^j > 0$. Thus there are incentives for the skilled to migrate from the rich to the poor country. This is not always the case, for example in Figure 8, corresponding to a rich country with a relatively high value of δ, $\lambda^i > 0$, but there is still full use of the modern technology. But differentials favouring skilled migration to the poor will certainly exist the greater the difference in fundamentals between rich and poor countries, i.e. for extreme δ, ρ differences.

Clearly the empirical record does not suggest any evidence of such incentives in reality. As Easterly (2002 p.80) notes, the incidence of migration to the United States from poor countries is larger amongst the educated than the uneducated. Some of this obviously reflects US immigration restrictions, but removing such restrictions would simply increase the incidence among the uneducated without likely altering the fact of high migration among the educated. For some cases, such out-migration of the educated is dramatic: Easterly quotes a "conservative estimate" that, in the late 90's, 77 percent of those with university education in Guyana had migrated to the United States.

This type of criticism has been levelled previously at accounts of growth that have attributed significant roles to skills, see for example Romer's (1995) comments on Mankiw's (1995) and (1992) emphasis on formal training. In the next section, we show that this is much less likely to be a problem for on-the-job skills in a more realistic, multiple technology environment.

4.1 Multiple Technology Steps

We now consider the implications of increasing the number of modern technology levels that are available, and we enrich our specification of the learning technology. Specifically, we consider a world in which the modern technology can be characterized by various gradations, corresponding perhaps to the generations of (exogenous) technology improvement. Accordingly, assume now that, in addition to a traditional technology with a constant marginal product $\overline{w}$, there are a finite number, I, of modern technology production processes, with corresponding production functions $\theta_i F^i(S_i, U_i)$ for each $i \in I$. The functions are ordered so that a higher index corresponds to a more advanced modern technology; $\theta_{i+1} F^{i+1}(S, U) > \theta_i F^i(S, U)$ for all S, U. The technologies satisfy identical Inada and homogeneity assumptions as in the previous section.

Skills-acquisition processes are more complicated than in Section 2. We assume these occur according to the technology specific functions $\Omega^i(S_i, U_i)$ with $\Omega_1^i \geq 0$, $\Omega_2^i \leq 0$ and Ω homogeneous of degree zero. This specification implies that the probability of becoming skilled is increasing in the proportion skilled and decreasing in the number of unskilled working with you. Skills acquired through each technology level are cummulative. Thus, in order to produce output working as an unskilled worker in technology $i+1$ it is necessary that a worker first be skilled in technology i. Workers skilled only in a lower technology are not able to leap-frog the skills acquisition process of earlier technologies; they produce no output and do not acquire on-the-job skills. However, we do allow for a type of condensed skills acquisition to occur through direct observation. In particular, each technology i admits direct skills acquisition through a process summarized by the function $\Omega^{iL}\left(S_i + U_i, U_i^L\right)$. The notation U_i^L denotes the number of workers not skilled in technology $i-1$ placed into direct learning of technology i. A worker placed into direct learning produces no output since they are not expert enough to even work as unskilled there, but by observing ongoing production, the level of which is given by total employment in the technology, $S_i + U_i$, acquires skills with probability $\Omega^{iL}(\cdot)$. We assume the function $\Omega^{iL}(\cdot)$ is identical in form to $\Omega^i(\cdot)$.

The outcome of a decentralized competitive equilibrium will again correspond to the solution to the planner's problem, so we proceed to consider the latter. We analyze an economy with only two modern technologies; $I = 2$. It will then be argued that the results relating to migration incentives are reinforced in a multiple modern technology setting. We again posit a constant population of size unity with the same overlapping generations structure. Now S_1 denotes the number skilled in technology 1, and S_2 those in technology 2, the remaining $1 - S_1 - S_2$ are unskilled in either technology. Let β_L denote the proportion of the unskilled allocated to direct skills acquisition, so that the remaining proportion $1 - \beta_L$ work in either the old technology, receiving $\overline{w}$ or as unskilled in technology 1. Let $\beta_1$denote the proportion of those unskilled not learning directly who work as unskilled in the new technology; the total number of unskilled thus doing so is $\beta_1 (1 - \beta_L)(1 - S_1 - S_2)$. The remaining $(1 - \beta_1)(1 - \beta_L)(1 - S_1 - S_2)$ work in the old technology. Let β_2 denote the proportion of those skilled in technology 1 working as unskilled workers in technology 2, the remaining proportion work in technology 1 as skilled workers. Consequently, output generated by technology 1 is $\theta_1 F^1 ((1 - \beta_2) S_1, \beta_1 (1 - \beta_L)(1 - S_1 - S_2))$ and that by technology 2 is $\theta_2 F^2 (S_2, \beta_2 S_1)$.

As is the case with a single technology, the planner's problem amounts again to choosing the β_i so as to maximize discounted lifetime utility. The relevant tradeoff is between working in a technology where one is already skilled and being highly productive, or foregoing current productivity and working as unskilled in a superior technology in order to become skilled with it. Additionally here, the planner may choose to place unskilled workers into the direct learning of technology 1 rather than have them trained through on-the-job skills acquisition, through working with that technology. This direct learning is denoted by $\Omega^L (S_2 + \beta_2 S_1, \beta_L (1 - S_1 - S_2))$. Since there is no higher level technology than 2, the function $\Omega^L(\cdot)$ refers, without ambiguity, only to learning proficiency in technology 1 and thus becoming able to work as an unskilled worker in technology 2.

Since all production is consumed immediately, a similar decomposition into cohorts as in the previous section allows us to write the Hamiltonian for the planner's problem as:

$$\begin{aligned}
&\max_{\beta_1,\beta_2,\beta_L} H = \overline{w}(1 - S_1 - S_2)(1 - \beta_L)(1 - \beta_1) + \theta_1 F^1 ((1 - \beta_2) S_1, \beta_1 (1 - \beta_L)(1 - S_1 - S_2)) \\
&+\theta_2 F^2 (S_2, \beta_2 S_1) \\
&+\lambda^1 \left[\begin{array}{c} \Omega^1 ((1 - \beta_2) S_1, \beta_1 (1 - \beta_L)(1 - S_1 - S_2)) \beta_1 (1 - \beta_L)(1 - S_1 - S_2) + \\ \Omega^L (S_2 + \beta_2 S_1, \beta_L (1 - S_1 - S_2)) \beta_L (1 - S_1 - S_2) - \Omega^2 (S_2, \beta_2 S_1) \beta_2 S_1 - \delta S_1 \end{array} \right] \\
&+\lambda^2 \left[\Omega^2 (S_2, \beta_2 S_1) \beta_2 S_1 - \delta S_2 \right]
\end{aligned}$$

$s.t.\ 0 \leq \beta_i \leq 1$ for $i = 1, 2, L,$

The first order conditions are:

$$\theta_1 F_2^1(\cdot) - \overline{w} + \lambda^1 \left[\Omega_2^1(\cdot)(1 - \beta_L)(1 - S_1 - S_2)\beta_1 + \Omega^1(\cdot) \right] \left\{ \begin{array}{c} = 0 \text{ if } \beta_1 \in (0,1) \\ < 0 \text{ if } \beta_1 = 0 \\ > 0 \text{ if } \beta_1 = 1 \end{array} \right. \tag{20}$$

$$\theta_2 F_2^2(\cdot) - \theta_1 F_1^1(\cdot) + \lambda^1\left[\Omega_1^L(\cdot)\beta_L(1-S_1-S_2) - \Omega_1^1(\cdot)\beta_1(1-\beta_L)(1-S_1-S_2) - \Omega_2^2(\cdot)S_1\beta_2 - \Omega^2(\cdot)\right]$$
$$+\lambda^2\left[\Omega_2^2(\cdot)S_1\beta_2 + \Omega^2(\cdot)\right]\begin{cases} = 0 \text{ if } \beta_2 \in (0,1) \\ < 0 \text{ if } \beta_2 = 0 \\ > 0 \text{ if } \beta_2 = 1 \end{cases} \quad (21)$$

$$-\overline{w}(1-\beta_1) - \theta_1 F_2^1(\cdot)\beta_1$$
$$+\lambda^1\left[\Omega_2^L(\cdot)(1-S_1-S_2)\beta_L + \Omega^L(\cdot) - \Omega_2^1(\cdot)\beta_1^2(1-S_1-S_2)(1-\beta_L) - \Omega^1(\cdot)\beta_1\right]\begin{cases} = 0 \text{ if } \beta_L \in (0,1) \\ < 0 \text{ if } \beta_L = 0 \\ > 0 \text{ if } \beta_L = 1 \end{cases} \quad (22)$$

$$\theta_1 F_1^1(\cdot)(1-\beta_2) - \overline{w}(1-\beta_L)(1-\beta_1) - \beta_1(1-\beta_L)\theta_1 F_2^1(\cdot) + \theta_2 F_2^2(\cdot)\beta_2$$
$$+\lambda^1\left[\begin{array}{c} \left[\Omega_1^1(\cdot)(1-\beta_2) - \Omega_2^1(\cdot)\beta_1(1-\beta_L)\right]\beta_1(1-\beta_L)(1-S_1-S_2) \\ -\beta_1(1-\beta_L)\Omega^1(\cdot) - \Omega_2^2(\cdot)\beta_2^2 S_1 - \Omega^2(\cdot)\beta_2 \\ +\left(\Omega_1^L(\cdot)\beta_2 - \Omega_2^L(\cdot)\beta_L\right)\beta_L(1-S_1-S_2) - \beta_L\Omega^L(\cdot) - \delta \end{array}\right]$$
$$+\lambda^2\left[\Omega_2^2(\cdot)\beta_2^2 S_1 + \Omega^2(\cdot)\beta_2\right] = \rho\lambda^1 - \dot{\lambda}^1 \quad (23)$$

$$\theta_2 F_1^2(\cdot) - \overline{w}(1-\beta_L)(1-\beta_1) - \theta_1 F_2^1(\cdot)\beta_1(1-\beta_L)$$
$$+\lambda^1\left[\begin{array}{c} \left(\Omega_1^L(\cdot) - \Omega_2^L(\cdot)\beta_L\right)\beta_L(1-S_1-S_2) - \beta_L\Omega^L(\cdot) - \Omega_2^1(\cdot)\beta_1^2(1-\beta_L)^2(1-S_1-S_2) \\ -\Omega^1(\cdot)\beta_1(1-\beta_L) + \Omega_1^2(\cdot)\beta_2 S_1 \end{array}\right]$$
$$+\lambda^2\left[\Omega_1^2(\cdot)\beta_2 S_1 - \delta\right] = \rho\lambda^2 - \dot{\lambda}^2. \quad (24)$$

We will not consider the full transitional dynamics of an economy in this case. A full characterization is considerably more complicated than the case in Section 2. Instead we will sketch the conditions for an economy to be in a steady state where only the traditional technology is used, and also the sufficient conditions for an economy to be in a steady state where only the highest technology, technology 2 here, is used. We shall assume that technology in the two cases is identical, and shall instead derive these sufficient conditions as restrictions on the parameters ρ and δ. We will then interpret these two situations as corresponding to two contemporaneous economies, and compare migration incentives across them. We are particularly interested in understanding the incentives for migration of the skilled.

A traditional economy

A traditional economy never develops the skills required to use either of the modern technologies. In such an economy workers remain fully employed in the traditional sector receiving wage $\overline{w}$. Consequently, such an economy is characterized by $\beta_L = \beta_1 = \beta_2 = 0$ and steady state skill levels $S_1 = S_2 = 0$. This is analogous to the situation depicted in Figure 7. The $\dot{\lambda}^1 = 0$ locus intersects the $\lambda_{\beta_1} = 1$ locus and the economy remains at a point like A. Once again, for given values of θ_i this occurs for sufficiently high values of ρ and δ. Denote the values of these for the poor country by ρ^P and δ^P. Figure 11 reproduces this steady state in the bottom left hand corner, denoted by A. Let λ^P denote the shadow value of skill in this economy. It is also depicted in the figure.

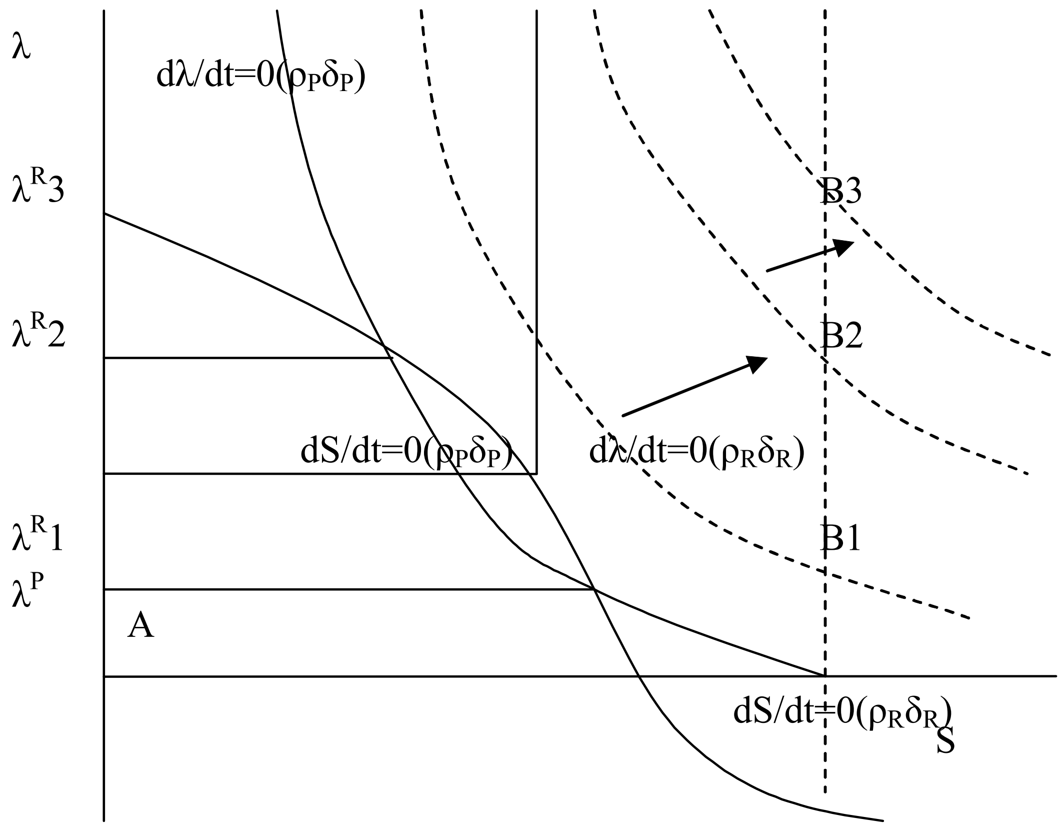


Figure 11

A fully modern economy

Now consider an economy that is fully specialized in the most advanced technology, technology 2. Recall that all individuals are born without skills in either technology 1 or 2, so that initially such individuals are not able to contribute to production. In a fully modern economy, new individuals are placed directly into the learning technology which will allow them to work as unskilled in technology 2. Production with Technology 1 does not occur. Once they attain the skills necessary to work as unskilled in technology 2, they productively enter the labor force. In this economy's steady state $\beta_L = \beta_1 = \beta_2 = 1$, S_1 and $S_2 > 0$, and since it is stationary

$\dot{S}_1 = \dot{S}_2 = 0$. The necessary conditions in this case are:

$$\theta_1 F_2^1(\cdot) - \overline{w} + \lambda^1 \Omega^1(\cdot) > 0 \qquad (25)$$

$$\theta_2 F_2^2(\cdot) - \theta_1 F_1^1(\cdot) + \lambda^1 \Omega_1^L(\cdot)(1 - S_1 - S_2) + (\lambda_2 - \lambda_1)\left[\Omega_2^2(\cdot) S_1 \beta_2 + \Omega^2(\cdot)\right] > 0 \qquad (26)$$

$$-\theta_1 F_2^1(\cdot) + \lambda^1\left[\Omega_2^L(\cdot)(1 - S_1 - S_2) + \Omega^L(\cdot) - \Omega^1(\cdot)\right] > 0 \qquad (27)$$

$$\theta_2 F_2^2(\cdot) + \lambda^1\left[\left(\Omega_1^L(\cdot) - \Omega_2^L(\cdot)\right)(1 - S_1 - S_2) - \Omega^L(\cdot) - \delta - \rho\right] + \left(\lambda^2 - \lambda^1\right)\left[\Omega_2^2(\cdot) S_1 + \Omega^2(\cdot)\right] = 0 \qquad (28)$$

$$\theta_2 F_1^2(\cdot) + \lambda^1\left[\left(\Omega_1^L(\cdot) - \Omega_2^L(\cdot)\right)(1 - S_1 - S_2) - \Omega^L(\cdot)\right] + \left(\lambda^2 - \lambda^1\right)\Omega_1^2(\cdot) S_1 - \lambda^2(\delta + \rho) = 0 \qquad (29)$$

$$\Omega^L(\cdot)(1 - S_1 - S_2) - \Omega^2(\cdot) S_1 - \delta S_1 = 0 \qquad (30)$$

$$\Omega^2(\cdot) S_1 - \delta S_2 = 0 \qquad (31)$$

Equations (28) to (31) can be solved for the four unknowns, $\lambda^{1*}, \lambda^{2*}, S_1^*, S_2^*$ in this steady state. By inspection it can be seen that conditions (30) and (31) uniquely determine S_1, S_2, and concavity of the technology then implies unique solutions to (28) and (29) for the co-state variables. If $S_2^* > S'$, defined from (2), then $\theta_2 F_2^2(\cdot) = \theta_2 F_1^2(\cdot)$ and $\lambda^{1*} = \lambda^{2*}$. We restrict analysis to this case in what follows, though qualitatively similar results follow if the steady states were not to have full abundance of those skilled in technology 2. In this case, equations (28) and (29) are the same, and the system remains determined with a unique solution to the remaining three variables.

Conditions (25) to (27) must also hold, but note that these involve the marginal product of workers (skilled and unskilled) placed in technology 1. However, with Inada restrictions, the productivity of a single skilled worker in this technology is undefined along the steady state path. Computing the value of a deviation by, for example, a skilled worker in technology 1 who, instead of working as unskilled with 2, works as skilled with 1, requires also solving for the induced accompanying unskilled in 1 that would accompany this worker. By doing this we obtain:

Lemma 5: *For sufficiently high productivity of the leading technology ($\theta_2 \gg \theta_1$) there exists a steady state in which only the leading technology is used.*

The steady state corresponding to this case can be partly depicted in the diagram already used. Since $\lambda^2 = \lambda^1$ we depict the relevant $\dot{\lambda}^1 = 0$ and $\dot{S} = 0$ functions in the same Figure 11 by dashed lines. The steady states are denoted in the right hand part of the figure by Bs. Where each one corresponds to a differing value of the productivity parameter, θ_2. The shadow value of skills is denoted λ^R in this case, and the figure denotes by $\lambda^R i$ the shadow value of skilled accompanying the rich country steady state Bi. In such steady states, the value of λ^R is determined from equation (29) as

$$\lambda^R = \frac{\theta_2 F_1^2(\cdot)}{\left(\Omega_1^L(\cdot) - \Omega_2^L(\cdot)\right)(1 - S_1 - S_2) - \Omega^L(\cdot) - \delta^R - \rho^R}.$$

Note that the arguments of Ω^L, S_1, S_2, are entirely determined by equations (30) and (31) which are independent of θ_2. Thus λ^R is monotonically increasing in θ_2. As shown in the diagram, increases in θ_2 correspond to upward shifts in the $\dot{\lambda}^1 = 0$ function.

Migration incentives:

We are now able to use Figure 11 to examine migration incentives in the modified, multiple technology model. Increases in θ_2 raise the value of λ^R. Consider then incentives for migration between the poor and rich countries. The simplest way to do this is to imagine successively higher values of θ_2 while making sure that the poor country remains in the traditional steady state, i.e. increasing ρ^P and or δ^P simultaneously so that the shadow value λ^P remains unchanged. For sufficiently high values of θ^2, individuals who are skilled in technology 2, which is abundant in the rich country and therefore receives no excess returns, will not wish to migrate to the poor country. This is because returns to working with the high productivity technology in the rich country, even as an unskilled worker, outweigh those to working with scarce skills in the poor country. This implies that the shadow value of being able to work as an *unskilled* worker in the high technology used in the rich country, λ^R, will not approach zero, even if the extra returns to being a *skilled* worker in that technology are negligible. Thus, the skill premium in the rich country's active technology can be zero, and hence the observed skill premium is also zero, while simultaneously being positive in the poor country without generating incentives for anyone in the rich country to migrate to the poor. We summarize with the following proposition:

Proposition 2: *If the leading edge technology is sufficiently productive, $\theta_2 \gg \theta_1$, then: (i) skilled workers do not want to move to poor countries and (ii) both skilled and unskilled workers in the poor country would prefer to move to the rich.*

Notice that this result will be even more pronounced the more layers of technology exist. In that case, individuals skilled in technology 2 will receive a skill premium upon migrating to the poor country, but will necessarily work with a far inferior technology there, and are even more likely still to have a lower marginal product.

5 Conclusion

The previous sections have demonstrated the following: on-the-job skills can generate arbitrarily large productivity increases without the existence of any externality to human capital, and when fully embodied in labor. Such skills can be an underlying source of productivity differentials across countries without implying any equilibrium skill premium. Skills accumulated in this way imply migration incentives that are consistent with those observed in reality - both skilled and unskilled migrate to rich countries

We have maintained the assumption throughout that competitive labor markets can adequately facilitate on-the-job skills acquisition. There are at least two reasons to doubt this in reality: labor is liquidity constrained and firms cannot write long-term contracts with labor.

As we saw, in transition, there are skill premia, $\lambda > 0$. Unskilled labor accept low wages, F_2, in order to become skilled and receive higher wages F_1 in future. However, especially at the start of this process, the net present value of 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 their skills acquired 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 for the employer at wages below market clearing when skilled. 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 whole discussion assumes away the possibility of worker moral-hazard, which is likely to become particularly severe when workers are being paid below market wage 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 skills acquisition does play a central role in productivity improvement, then it should be the case that countries that have experienced growth miracles 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. Or, institutions that allow labor to overcome these problems on its own, and thus become skilled. There is certainly evidence of such institutions in some of the miracle economies. For example, Japan and South Korea, 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.

A full survey of these literatures is beyond the scope of this paper, but further examination of such cases is certainly motivated by the findings here. The implied research direction is almost the converse to, and perhaps complementary with, a "barriers to adoption" view of development, as suggested by Parente and Prescott (2002). A barriers view would argue that because incentives for technological upgrading are so huge in decentralized economies, large institutional impediments must exist to thwart otherwise productive undertakings. The implication is that if governments, in particular, can be convinced to remove these impediments, the flow of ideas will be freed and productivity in LDCs will rise. In contrast, our framework suggests that, even for economies without institutional impediments and with full and free access to the modern technology, productivity improvements may not be forthcoming. This may even be the case for countries able to maintain such sustained improvements if factor markets do not well approximate the competitive paradigm. One interpretation of countries stagnating with low productivity growth then is that, far from being saddled with institutional burdens to development that strangle development, these countries may simply not yet have developed the relatively complex institutions needed to encourage sustained skills accumulation in the presence of factor market imperfections - like contracting incompleteness and liquidity constraints. From this perspective, it is certainly not surprising that many countries (especially in Sub-Saharan Africa) that have responded diligently to decentralizing structural reforms that were inspired by a competitive view of factor markets, have not been able to step onto the escalator of rapidly increasing productivity growth.

References

[1] Aghion, P. and P. Howitt (1992) Aghion and Howitt (1992) “A Model of Growth Through Creative Destruction” Econometrica, 60, 323-351.

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

[3] 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.

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

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

[6] 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

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

[8] 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.

[9] 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.

[10] 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.

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

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

[13] 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.

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

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

[16] Pack and Paxson (199?)

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

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

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

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

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

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

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

[24]

[25]

Appendix

Proof of lemma 1:

Suppose the contrary, then from (12) necessarily $w_t^u \leq \overline{w}$. $\overline{\gamma} > 0$ also implies $\gamma^s < 1$ so from (11) $w_t^s \leq w_t^u$. But $w_t^s < w_t^u$ can be ruled out directly since from (11) this implies $\gamma = 0$ so that it implies $w_t^s \equiv \theta F_1 (0, \cdot) < \theta F_2 (0, \cdot) = w_t^u$, which violates concavity of the technology. Thus the only remaining possible implication is $w_t^s = w_t^u \leq \overline{w}$. But $w_t^s = w_t^u$ implies $F_1 = F_2$ so that from condition (3) we have $\theta F_2 \left(S', U'\right) = w_t^u > \overline{w}$. Which is a contradiction ■

Proof of lemma 2:

Since $\overline{w}$ and θ are constant, the minimization rearranges to the simpler: $\min_S \Omega F_1 (\cdot) + (\delta + \rho) F_2 (\cdot)$. The first order condition is: $\Omega (F_{11} - F_{12}) + (\delta + \rho) (F_{21} - F_{22}) = 0$. Since $F_i (\cdot)$ are homogeneous of degree 0, we know $SF_{11} + (1 - S) F_{12} = 0$ and $SF_{21} + (1 - S) F_{22} = 0$, which imply after rearrangement: $F_{11} - F_{12} = -\frac{F_{12}}{S}$, and $F_{22} - F_{21} = -\frac{F_{21}}{(1-S)}$. Substituting in to the first order condition yields: $-\Omega \frac{F_{12}}{S} + (\delta + \rho) \frac{F_{21}}{1-S} = 0$, which, since $F_{21} = F_{12}$ imply the unique solution is $S = \frac{\Omega}{\Omega + \delta + \rho}$. Second order conditions confirm this is a minimum.■

Proof of lemma 3:

When $\beta < 1$, $\dot{\lambda} = 0$ implies $\overline{w} = F_1 (S, \beta (1 - S)) - (\delta + \rho)\lambda$. Differentiating this condition with respect to S yields: $0 = F_{11} + F_{12} \left(\frac{d\beta}{dS} (1 - S) - \beta\right) - (\delta + \rho) \frac{d\lambda}{dS}$. $\beta < 1$ implies $\overline{w} = F_2 (S, \beta (1 - S)) + \Omega\lambda$. Differentiating this with respect to S yields $0 = F_{21} + F_{22} \left(\frac{d\beta}{dS} (1 - S) - \beta\right) + \Omega \frac{d\lambda}{dS}$. Substituting out the β terms and rearranging these two differentiated equations yields: $\frac{d\lambda}{dS} \left(\delta + \rho + \frac{F_{12}}{F_{22}} \Omega\right) = F_{11} - \frac{F_{12} F_{21}}{F_{22}}$. But as $F (\cdot)$ is homogeneous of degree 1, we have $\frac{F_{11}}{F_{12}} = \frac{F_{21}}{F_{22}}$, implying that $\frac{d\lambda}{dS} = 0$.■

Proof of lemma 4:

$\dot{S} = 0$ in $\beta < 1$ phase implies $\Omega (1 - S) \beta = \delta S$. Differentiating with respect to S yields $\frac{d\beta}{dS} (1 - S) - \beta = \frac{\delta}{\Omega}$. From the previous proof we have that the corresponding term obtained from the $\beta < 1$ condition is $0 = F_{21} + F_{22} \left(\frac{d\beta}{dS} (1 - S) - \beta\right) + \Omega \frac{d\lambda}{dS}$. Substituting for the β terms and rearranging yields $\frac{d\lambda}{dS} = \frac{-F_{21} - F_{22} \frac{\delta}{\Omega}}{\Omega}$. Since $F_2 (S, \beta (1 - S))$ is homogeneous of degree 0 we know that $SF_{21} + \beta (1 - S) F_{22} = 0$, and $\dot{S} = 0$ implies $(1 - S) \beta = \frac{\delta}{\Omega} S$. Combining these two conditions yields $F_{21} + \frac{\delta}{\Omega} F_{22} = 0$. But this implies $\frac{d\lambda}{dS} = 0$.■

Proof of Proposition 1.

Proof of part a: The $\dot{\lambda} = 0$ crosses the $\lambda_{\beta=1}$ locus when, for a value of S, $\frac{\overline{w} - \theta F_2}{\Omega} - \frac{\theta (F_1 - F_2)}{\delta + \Omega + \rho} > 0$. From Lemma 2 a unique value of S minimizes the distance between these two terms. Consequently, for given θ, a unique value k^* solves: $\frac{\overline{w} - \theta F_2}{\Omega} = \frac{\theta (F_1 - F_2)}{k^* + \Omega}$. If and only if $\rho + \delta < k^*$ the functions do not cross and situations in Figures 6 or 8 occur, so that, in steady state, no one works in the traditional technology.■

Proof of part b: S' is defined from equation (2). If and only if $\frac{\Omega}{\Omega+\delta} > S'$ 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 6 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', (Figures 6,7 or 8) the steady state involves $\lambda>0$. Thus $\delta^*=\frac{\Omega(1-S')}{S'}$.

Proof of part c: δ^{**} solves $\theta F_2\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)=\overline{w}$. 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 no intersection with the $\dot{\lambda}=0$ locus and the unique steady state is in the interior, as in Figure 6. If $\delta+\rho\geq k^*$, there is an interesection and the case of Figure 9 applies, implying hysteresis. For $\delta\geq\delta^{**}$, the $\dot{S}=0$ locus intersects the $\lambda_{\beta=1}$ line. Whether the economy is in the hysteresis case, or the stable 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:

$$\frac{\theta F_1\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)-\theta\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)}{\rho+\Omega+\delta}=\frac{\overline{w}-\theta F_2\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)}{\Omega}$$

$$\implies \rho^*=\frac{\Omega\theta F_1\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)+\delta\theta F_2\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)}{\overline{w}-\theta F_2\left(\frac{\Omega}{\Omega+\delta},\frac{\delta}{\Omega+\delta}\right)}.$$

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 7 occurs, and the unique steady state is full use of the traditional technology. Under the converse, the situation in 9 occurs and hysteresis is the outcome.■

Proof of Lemma 5:

Find values λ^{i*}, S_i^* satisfying (28) to (31). For these values, consider a deviation by an unskilled worker in technology 2 to skilled in technology 1 to verify that condition (26) holds. This condition requires computing $\theta_1 F_1^1(\cdot)$. Note that, under such a deviation, necessarily, those unskilled in technology 1 are allocated according to (20) with $0<\beta_1<1$. Thus we have:

$$\theta_1 F_2^1(\cdot)=\lambda^1\left[\Omega^L(\cdot)+\Omega_2^L(\cdot)-\Omega^1(\cdot)\right].$$

This equation yields a solution for the ratio of skilled to unskilled working in technology 1, denote this x. Importantly, x is independent of θ_2. It also follows from the linear homogeneity of F that, under the deviation:

$$\theta_1 F_1^1(\cdot)=\frac{\theta_1 F^1(\cdot)-\theta_1 F_2^1(\cdot)}{x}.$$

Substituting both of these terms into condition (26) yields the restriction:

$$\theta_2 F_2^2 (\cdot) - \frac{\theta_1 F^1 (\cdot) - \lambda^1 \left[\Omega^L (\cdot) + \Omega_2^L (\cdot) - \Omega^1 (\cdot)\right]}{x} + \lambda^1 \Omega_1^L (\cdot) (1 - S_1 - S_2) > 0,$$

which holds for θ_2 sufficiently high.■