A Theory of Technology Diffusion

Toshihiko Mukoyama*
Department of Economics
Concordia University
and
CIREQ

April 2003

Abstract

What determines the speed of the technology diffusion? What are the consequences of diffusion? This paper presents a model to address these questions. Skilled machine-users adopt a new technology first, while unskilled users wait until machines become more reliable and accessible. The quality improvement of machines is the engine of diffusion, and it is carried out by the machine producer. The speed of diffusion is affected by the skill distribution in the economy. At any point in time, the machine producer can start producing a new generation of machines. The timing of this event is influenced by the skill distribution.

Keywords: Technology Diffusion; Skill; Quality Improvement; Learning by Using; R&D; Innovation

JEL Classification: J31; O31; O33

---

*This is a revised version of Chapter 1 of my Ph.D. dissertation submitted to the University of Rochester. I would like to thank Jeremy Greenwood for his constant guidance and support. This paper has benefited from comments of Mark Bils, Jean-Pierre Danthine, Samuel Danthine, Po-Han Fong, Per Krusell, Josef Perktold, and seminar participants at Clemson University, Concordia University, Cornell University, Queen's University, University of Pittsburgh, University of Rochester, and York University. It has also benefited from conversations with Daron Acemoglu, Stanley Engerman, Shakeeb Khan, and Walter Oi. All remaining errors are my own.

1 Introduction

New technologies are the engines of economic growth. Since the pioneering work of Abramovitz (1956) and Solow (1957), many researchers have found that technological progress is essential in growth process.[1] Recent studies on international income differences suggest that a large part of income variation is explained by the differences in technology employed in each country.[2]

It takes time for a new technology to acquire economic significance. First, it has to be brought into the economy (*innovation*). Then, it is gradually adopted by many people (*diffusion*). The last decade has witnessed a large development in the economics of innovation. However, not much attention has been paid to the economics of diffusion. Diffusion is as important as innovation: no new technologies have an economic impact until they become widespread in the economy. Diffusion is not a trivial process: in general, diffusion takes a long period of time.[3] Moreover, in many cases, innovation and diffusion are interrelated. This paper is an attempt to understand the process of technology diffusion.

A common feature of newly invented machines is that initially they are difficult to handle. This feature leads to a well-known empirical fact: high levels of skill[4] are required in the early stage of technology diffusion. To quote Nelson, Peck, and Kalachek (1967),

> The early ranks of computer programmers included a high proportion of Ph.D. mathematicians; today, high school graduates are being hired. During the early stage of transistors chemical engineers were required to constantly supervise the vats where crystals were grown. As processes were perfected, they were replaced by workers with less education. (pp.144–145. As quoted by Bartel and Lichtenberg 1987)

[1] Taking the period from 1929 to 1982, Denison (1985, p.30) estimates that 34% of the growth in U.S. real nonresidential business output (after smoothing business cycles) comes from "the advances in knowledge". Barro and Sala-i-Martin (1995, Table 10.8) summarize some results on other countries.

[2] See, for example, Klenow and Rodríguez-Clare (1997).

[3] Mansfield (1968) reports: "Measuring from the date of the first successful commercial application, it took twenty years or more for all the major firms to install centralized traffic control, car retarders, by-product coke ovens, and continuous annealing" (p.115).

[4] By skill we mean higher level of ability and education. We especially focus on a specific aspect of skill: the ability to cope with new technologies. In a broad sense, it corresponds to what Schultz (1975) called "allocative ability" – the ability to adjust to the changes in economic conditions. Rogers (1995) summarizes the characteristics of early adopters of innovations. Among them, it is described that: "early adopters have more years of formal education than later adopters" and "early adopters are more likely to be literate than later adopters" (p.269).

This is the first key ingredient of our theory. The idea that skill is required for the adoption of a new technology is also expressed by other economists. In an influential paper, Nelson and Phelps (1966) construct models where education enhances adoption. They obtain insightful results such as: the return to education is higher if the technological progress is faster, and the level of technology adopted is higher when the level of education is higher. These results are consistent with our model. However, Nelson and Phelps do not consider the *process* of technology diffusion explicitly. In their model, the law of motion for adoption is given exogenously, while we make the process of diffusion endogenous.

To investigate the endogenous process of technology diffusion, we ask the following question: if a new technology requires skill, how is it possible that it will eventually be adopted by less-skilled users? Casual observations suggest that it is due to a change in the nature of the technology. As the technology matures, machines become more user-friendly so that anyone can handle them. They also become more reliable. Figure 1 is taken from Rosenberg (1982, p.131). It exhibits the typical cost of an engine maintenance schedule, together with the particular trajectory for the Pratt and Whitney JT3D turbojet engine. It shows that maintenance expense drops dramatically in the first several years of engine operation. This reflects improvement in the reliability of engines. The increase in reliability makes the adoption easier for less-skilled adopters. This change in nature of technology is the second key ingredient of our theory.

How does this change happen? We observe that it is mainly achieved by an effort to improve the technology by the producers of machines. The producers of a new machine can improve the machine (improve reliability and user-friendliness) by learning and investing in R&D. This aspect has been analyzed previously. Stoneman and Ireland (1984) is an early attempt to model the supply side effect of diffusion. They consider a capital good monopolist facing users with different threshold price levels. The production cost falls by learning by doing, and the price of a machine falls over time. This leads to the diffusion of the technology. Unlike our interest in skill and improvement, their focus is on the price decline, and relating the price decline to the shape of the diffusion curve.

In this paper, a model that focuses on supply-demand interaction in the process of technology diffusion is constructed. On the demand side, the skill levels of the users are heterogenous, and high-skilled users adopt the technology early on. On the supply side, improvement of technology occurs through the producer's learning and R&D activity. It makes the adoption by less-skilled users possible. We also allow a producer to start producing a new generation of machines at any time.

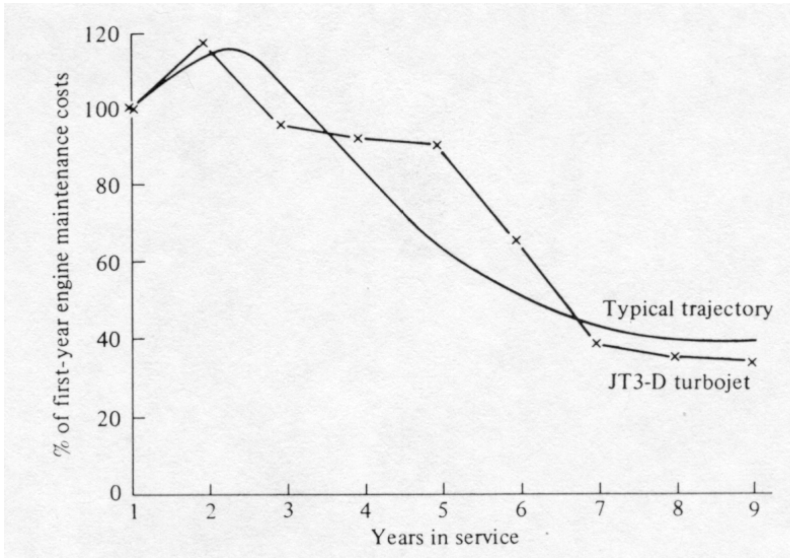


Figure 1: Engine Maintenance Expense

Empirical Background: Many recent empirical studies examine the relationship between the adoption of new technologies and the skill level. Bartel and Lichtenberg (1987) show that there is a positive correlation between skill and new technology adoption. They consider the hypothesis that a firm that adopts new technologies demands more skilled workers. To test their hypothesis, they use the share of skilled workers as the dependent variable and the age of the capital as an independent variable. In other words, they consider that high skill level is the *result* of new technology adoption. We have a different causality in mind: a firm with more skilled workers is more likely to adopt a new technology.[5] A recent study by Doms, Dunne, and Troske (1997) suggests that this direction is more relevant. They conduct two empirical analyses: cross-sectional and time-series. In the cross-sectional analysis, they find a positive correlation between skill level (measured by education or wage level) and technology use. In the time-series analysis, they did not find any change in the worker characteristics between before and after the adoption of new technologies. They interpret the result as:

[5] The source of the difference seems to be the view about which is more fixed (and fundamental) for a firm; skilled workers or technology. Bartel and Lichtenberg consider a story where new machines are fixed and they call for skilled workers, and we view that skilled workers are fixed resources for a firm and they adopt new machines.

Attribute	5 (Very Important)	4	3	2	1 (Not Important)	Mean
Quality/reliability of product	79%	19%	2%	0%	0%	4.8
Ease of use	43%	45%	10%	1%	0%	4.3
Price/performance	48%	37%	13%	1%	0%	4.3
Service/support	43%	44%	11%	1%	1%	4.3
Ease of maintenance	42%	41%	16%	2%	0%	4.2
Purchase price	37%	45%	14%	4%	1%	4.1

Table 1: Importance of Attributes When Evaluating Computers, Computer Boards and Peripherals Vendors and Products

> Our findings suggest that, at the plant level, the correlation between technology use and worker wages is primarily due to the fact that plants with high wage workforces are more likely to adopt new technologies. (p.255)

We follow Doms, Dunne, and Troske and consider skill as a *determinant* (not a result) of a new technology adoption. More recently, Caselli and Coleman (2001) analyze the diffusion of computers among countries. They show that the human capital level in each country is a significant determinant of computer imports.

We emphasize the importance of a machine's reliability and ease of use in adoption decision. Table 1 is taken from Cahners Electronics Group (2000). It is based on a survey of north American design and development engineers. It shows that the factors such as quality/reliability and ease of use are more important than price/performance when they evaluate computers, computer boards and peripherals.

The role of capital goods producers in the process of the technology improvement is emphasized by many economic historians. For example, MacLeod (1992) studies mechanical engineering industry in 19th century Britain and writes "... it was often only through the medium of their capital-goods suppliers that information about a new technology was passed back and forth among users" (p.287). In our model, this channel is important since the experience of current users contributes to the improvement of the technology and benefits future adopters only through the supplier.

Recent Literature: The role of the change in the nature of technology during diffusion tends to be ignored in the recent literature. One important exception is Galor and Tsiddon (1997). They explicitly analyze the effect of the improvement in accessibility of a technology on wage inequality, intergenerational earnings mobility, and economic growth. They take the improvement in accessibility as an *exogenous* process, and consequently do not analyze the mechanism of the improvement itself. In contrast, in this paper the process of improving a technology, which is an essential determinant of the technology diffusion, is modelled *endogenously.*

There are several important recent papers that attempt to explain diffusion of new technologies as endogenous process. Jovanovic and MacDonald (1994) analyze innovation and diffusion of knowledge. Firms try to acquire better knowledge (technologies) by R&D and learning. Diffusion of new technologies is slow because of informational barriers: it takes time and effort for firms to learn new technologies. Technologies are disembodied, and there is no role for capital goods producers. (In fact, there is no capital in their model.) In our model, the capital goods producer plays a crucial role in both innovation and diffusion. We do not assume any informational barriers, but lack of skill prevent some people from using a new machine. Chari and Hopenhayn (1991) construct a vintage human capital model where new and old capital are complementary inputs. The marginal product of investment depends not only on the vintage of technology but also on the amount of old capital available for that specific vintage. Even when new technologies are available, people invest in old technologies if there exist abundant old capital for these technologies. As a consequence, diffusion of new technologies is slow. We do not assume any complementarity between old and new capital. In our model, diffusion is slow because new machines are difficult to use. Jovanovic and Lach (1989) study diffusion of a new technology which is embodied in capital goods. The vintage-specific installation cost and production cost fall over time, due to (external) learning by doing. This leads to a gradual diffusion of a new technology. Thus, the diffusion is totally due to the fall of the adoption cost. Their main focus is on the shape of the diffusion curve. In their model, firms are homogeneous before adoption; in contrast, we emphasize the skill difference among machine users.

Since the rate of the technological progress is determined endogenously, this paper is also related to the literature on endogenous technological progress and growth. The most relevant model is Young (1993). Young constructs an endogenous growth model with innovation and learning by doing. In his model, a new product is invented by R&D activity. A new product has high production cost initially, but it falls as learning by doing occurs. His model focuses on cost decline through learning in the consumption goods sector. In contrast, in our model the learning occurs in capital goods sector, and learning induces an improvement of a technology. His model does not exhibit gradual diffusion: consumers are homogenous and a new product is either purchased by everyone or not purchased at all.

In our model, machines become more and more accessible by learning and R&D investment. For learning, we consider a specific mechanism: *learning by using.* A statistical model is employed to derive a learning function (Lemma 1). Jovanovic and Nyarko (1995) derived a different form of learning function based on Bayesian theory. They consider a situation where a worker tries to find an unknown target value through trial and error. In contrast, we study a setting where the capital goods producer learns from the experience of users.

Structure of the Paper: In the next section, a model of technology diffusion is constructed. In Section 3, the dynamics of the model is analyzed through numerical simulations. In Section 4, the effect of the change in skill distribution is analyzed. In Section 5, some other implications and possible extensions of the model are discussed. Section 6 concludes.

2 Model

Consider a capital good industry, where machines are sold to end-users. The market is monopolized by one producer. For simplicity, it is assumed that the monopoly lasts forever and there is no entry into the machine production.

A machine has two dimensions of characteristics: *performance* and *quality.* Performance is the measure of the benefit that one can obtain from using the machine. *Machines with higher performance* are, for example, more efficient, more powerful, faster, and so on. Quality corresponds to the characteristics such as likeliness to break down, ease of use, frequency of maintenance, and so on. *Higher quality machines* are more accessible and easier to adopt. In the model, quality is represented by the error-free rate: if a machine contains many errors, it is called a low-quality machine. Later, we impose an assumption that a low-quality machine can be operated only by high-skilled users. As the quality gets better, lower-skilled users become able to adopt it. In this sense, quality substitutes for the skill of users.[6]

[6] Quality improvement in our model can be interpreted as a "skill-replacing" technological progress. For discussions that the technological advances were predominantly skill-replacing in the nineteenth century, see Acemoglu (2002) and Goldin and Katz (1998).

Performance is denoted by $b_i \in \mathbb{R}_{++}$, where i is the index of the generation of the technology. When the generation-i technology is born, the value of b_i is determined. b_i does not change over the life cycle of technology i. It is always the case that $b_i \geq b_{i-1}$ for all i. Quality is represented by the error ratio, $\theta_i \in [0, 1]$. It represents the amount of "bugs" in a generation-i machine. Smaller θ_i implies higher quality. θ_i is improved over time as technology i matures. (Thus, θ_i is inversely related to the quality. We can view $(1 - \theta_i)$ as the actual measure of the quality.) It is possible that $\theta_{i-1} < \theta_i$ (an older technology has higher quality), especially when i is just introduced.[7] Thus, there is a possibility of a trade-off between performance and quality. In equilibrium, skilled users choose to adopt a higher-performance technology despite its low quality, while unskilled users remain using a lower-performance technology because of its high quality.

2.1 Machine Users

A user of a capital good (machine) is defined as a worker-firm match. Each worker has an exogenously given skill level s. This is going to be a key factor in the technology adoption decision. The machine user produces a consumption good. The only input for the consumption good production is a machine (and its operator). The machine depreciates completely in one period. Let b_i be the amount of surplus generated by hiring one generation-i machine.

Imagine that, when introduced, a machine contains "bugs" when performing some of its tasks. Let θ_i be the ratio of the tasks that contain bugs in a generation-i machine. The error rate, θ_i, declines over time as bugs are corrected. This process is called a *quality improvement*. Assume that high-skilled users are more tolerant of the errors. Low-skilled users and high-skilled users are hit by an error with the same probability, but high-skilled users can deal with the bugs better. They are able to solve the technical problems and fix the bugs more easily.[8] The profit of a user from

[7] Mansfield (1968) states: "Learning takes place among the producers of innovation, as well as the users. Early versions of an innovation often have serious technological problems, and it takes time to work out these bugs. During the early stages of the diffusion process, the improvements in the new process or product may be almost as important as the new idea itself" (pp.112–113).

[8] Anderson and Davidson (1940) describe: "In the first stages of such mechanization when machines are somewhat crude, their operation requires the watchful care of skilled machinists. Likewise, to make those improvements which increase machine efficiency, competent machinists are employed to observe the machines in operation under practical working conditions. But, as machines increase in importance, they must be further improved in efficiency so as to require little attention and a minimum number of stoppages for repairs or overhauling. This increasing efficiency of the machine itself tends, in the long run, to eliminate much of the work of that large corps of machinists which was required when machines were first installed to displace hand workers" (p.228. As quoted by Goldin and Katz 1998).

using technology i ($i = 0, 1, 2, ...$) is assumed to be

$$\pi_i(s, b_i, \theta_i) = b_i - g_i(s, \theta_i), \tag{1}$$

where $g_i(s, \theta_i)$ represents the cost of handling errors. The function $g_i(s, \theta_i)$ is decreasing in s and increasing in θ_i. We consider an extreme case: assume that $g_i(s, \theta_i) > b_i$ if $s < m\theta_i$ and $g_i(s, \theta_i) = 0$ otherwise, where $m > 0$ is a constant parameter. Then $\pi_i(s, b_i, \theta_i) \geq 0$ if and only if $s \geq m\theta_i$. A user can earn positive profit if and only if he possesses large enough skill level relative to θ_i. Thus the *potential* demand for a machine with technology i and error rate θ_i, when it is priced at p_i, is[9]

$$D_i(p_i, \theta_i) = \begin{cases} 0 & \text{if } p_i > b_i, \\ N[1 - F(m\theta_i)] & \text{if } p_i \leq b_i, \end{cases} \tag{2}$$

where $F(\cdot)$ is the distribution function of skill. $D_i(p_i, \theta_i)$ is called *potential* demand, since all technologies are subject to the competition with other technologies. Thus, *actual* demand in equilibrium can be smaller than $D_i(p_i, \theta_i)$.

One advantage of this formulation is that we can conceptually distinguish between the *quality improvement of the existing technology* (a fall in θ_i) and the *arrival of a new technology* (an increase in b_i) in a simple fashion. The current demand specification contrasts them in an extreme way: the quality improvement expands the demand (shifts the demand curve rightward), while an innovation that improves performance increases the reservation price (shifts the demand curve upward). See Figure 2. It also allows for the effect of a quality improvement in technology to be separated from the effect of a decline in machine production cost.[10] In the current model, the production cost does not affect demand as long as the marginal cost is lower than b_i.

2.2 Machine Producer

2.2.1 Production

The monopolist can produce the current ith-generation technology and the past $(i-1)$th one. For simplicity, the production cost of the machine is set to zero.

[9] Alternatively, this form of demand can be interpreted as reflecting a *skill requirement* for the use of a machine with technology i (as θ_i gets smaller, the machine becomes easier to use).

[10] As discussed before, the implications of machine production cost changes are already analyzed in the existing literature (see, for example, Stoneman and Ireland 1984 and Jovanovic and Lach 1989).

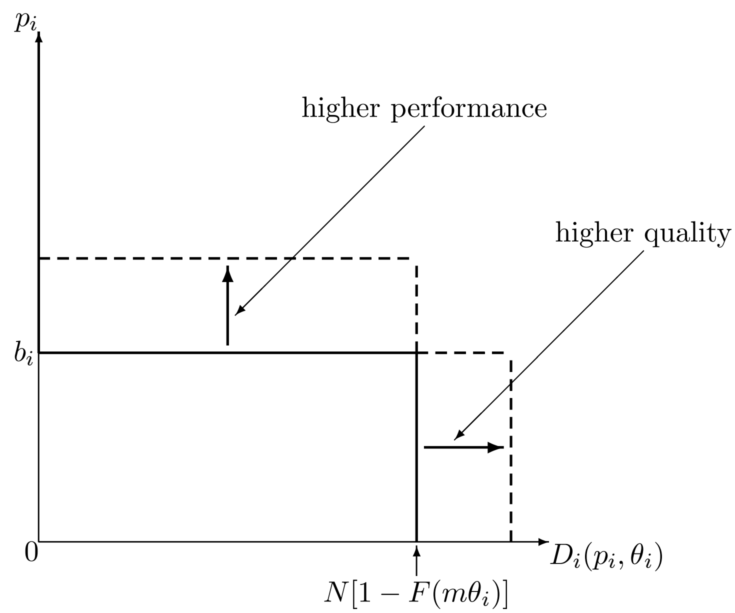


Figure 2: Potential Demand

2.2.2 Innovation

He can always begin producing newer $(i+1)$th-generation machines, which provide a higher performance, b_{i+1}. We call this switch *innovation*. Upon innovation, two things happen:

1. The monopolist has to abandon the technologies $(i-1)$ and older. He is able to maintain only two production lines, i and $(i+1)$, at the same time.

2. The $(i+1)$th-generation technology starts from the quality of

$$\theta_{i+1} = \theta_i + c(b_{i+1}/b_i),$$

   where $c(\cdot) : [1, \infty) \to \mathbb{R}_+$ is increasing and continuous. Quality is lost by upgrading, and the amount of the lost quality is larger when the producer introduces the higher performance technology.[11]

[11] A more general specification is

$$\theta_{i+1} = \nu \cdot \theta_i + c(b_{i+1}/b_i),$$

where $\nu \in [0, 1]$. When $\nu < 1$, there is an *incomplete transfer* of quality. It turns out that, in the numerical analysis, the qualitative nature of the model is not altered by the value of ν.

The first assumption reflects the capacity constraint in production. It is easy to extend two lines to more lines, without changing the qualitative nature of the model.[12] It ensures that, in equilibrium, there exist (at most) two kinds of active users: the ones who operate the ith-generation technology and the ones who employ the $(i-1)$th-generation technology.

The second assumption bears some similarity to the vintage capital literature (e.g. Zeckhauser 1968 and Parente 1994). A large b_{i+1}/b_i implies that the nature of technology $i+1$ is very different from technology i, and the knowledge that the machine producer has accumulated for technology i cannot be applied to technology $i+1$. It introduces a trade-off in the choice of innovation. Even though higher b_{i+1} can give a higher unit profit to the monopolist, it increases θ_{i+1} so that the demand for the $(i+1)$th-generation technology becomes smaller. The value of θ_{i+1} is restricted to be less than one. From $\theta_i + c(b_{i+1}/b_i) \leq 1$, it follows that $b_{i+1} \leq b_i c^{-1}(1-\theta_i)$.

2.3 Quality Improvement

The technology can be improved over time by *learning* and *R&D investment*. Both activities by the machine producer contribute to the decline in θ.

2.3.1 Quality Improvement-1: Learning by Using

First, consider the process of learning. Here, a specific form of learning, *learning by using*, is considered. The capital goods producer learns from the experience of users.[13] For example, users may find bugs in the machine, which leads to improvement in machine quality. Users may be able to suggest new applications of the machine. Rosenberg (1982) stated:

> ... in an economy with complex new technologies, there are essential aspects of learning that are a function not of the experience involved in producing the product but of its utilization by the final user. ... Perhaps in most general terms, the performance characteristics of a durable capital good often cannot be understood until after prolonged experience with it. (p.122)

[12] Alternatively, we can introduce a cost of expanding product lines. Then, capacity is determined endogenously.

[13] In the literature of economic development, the effect called "learning by exporting" is often discussed to be important source of technology improvement. Grossman and Helpman (1991) notes that "When local goods are exported the foreign purchasing agents may suggest ways to improve the manufacturing process ..." (p.166). It is often argued that "learning by exporting" has been important in Korea's economic development. See Rhee, Ross-Larson, and Pursell (1984, Chapter 4).

In what follows, we will employ a statistical model to formulate the idea of learning by using.[14]

Learning Process Learning occurs on the side of the machine producer.[15] As discussed above, learning by machine producers is emphasized in many studies in economic history. However, this aspect has not been formulated in theoretical literature.

Machine producers learn from the users of machines. Each machine can be used for K tasks. There are Z users who operate machines in each period. Each user employs his machine to carry out one of these tasks, chosen at random. The operation of $J \leq K$ tasks are subject to "bugs". Thus, a user is hit by an error (finds an error) with probability J/K. For the ease of exposition, suppose that the users operate machines in turn. If the first user finds an error, it is reported to the producer and fixed instantly, and the probability that the second user finds an error drops to $(J-1)/K$. If the first user does not find an error, the probability that the second user finds an error remains J/K. Then the third user operates the machine, and so on.

The following lemma summarizes the learning process.

Lemma 1 *When J, K, and Z are large, the expected ratio of remaining errors in a machine, θ', follows the law of motion*

$$\theta' = \theta \cdot e^{-z}, \tag{3}$$

where $\theta = J/K$ is the initial error ratio and $z = Z/K$.

[14] Conceptually, it is important to distinguish between *learning by using* and conventional *learning by doing* (Arrow 1962). We consider learning by using as a process of *quality improvement* through the *utilization* of the capital good. In contrast, learning by doing is usually formulated as the *decline of production cost* induced by the cumulative experience of *production*. In many cases the cumulative utilization and the cumulative production of a capital good move in the same direction (in our model they coincide), and it is difficult to formally differentiate between them (unless the capital goods are under-utilized or the capital-labor ratio changes over time). However, the difference between quality improvement and cost decline is evident in our specification. It will become clear that in equilibrium, quality improvement (decline of θ_i) enhances diffusion, while cost decline has no impact on diffusion (as long as the cost of production is lower than b_i).

[15] This assumption provides an interesting channel of spillovers. It is possible that skilled and unskilled workers in our model are producing different consumer products (for example, consider the case that the new capital good is a computer). The experience of a skilled worker may generate a positive spillover for an unskilled worker, if it makes it easier for the unskilled worker to adopt a new technology. Then, this can be a channel of inter-industry learning spillover. Stokey (1988) develops a growth model with inter-industry spillovers and writes "... it is important that learning display spillovers among goods. Otherwise, learning simply reinforces existing patterns of production ...". We do not explicitly model the production structure of the goods produced by the users, but an extension along the lines of Stokey (1988) would give a similar implication.

Proof. See Appendix. ■

Notice that z measures the number of users relative to the total number of parts. Although the above process of error finding is stochastic, for simplicity we posit a deterministic process for learning. Following the law of motion (3), we postulate a learning function[16]

$$\theta_i' = \theta_i \cdot e^{-\mu Y_i}, \tag{4}$$

where $\mu > 0$ is the parameter of learning efficiency and Y_i is the flow machine use of the generation-i technology.

2.3.2 Quality Improvement-2: R&D

It is assumed that the quality of a technology can also be improved through R&D investment. That is, the producer can make its product more user-friendly by "debugging" it. To do so requires a cost.

We assume that R&D cost is a function of the generation of technology and the amount of quality improvement. Specifically, it is assumed to take a form of $b_i \cdot R(x)$, where x is the amount of quality improvement. The cost is proportional to the performance. The function $R(\cdot)$ relates the amount of quality improvement to the cost, and satisfies that $R(\cdot) \geq 0$, $R'(\cdot) \geq 0$, and $R(0) = 0$.

[16] The learning process (4) has two attractive features. First, using one machine for two periods leads to the same degree of improvement as using two machines for one period. This transpires since $(\theta_i \cdot e^{-\mu}) \cdot e^{-\mu} = \theta_i \cdot e^{-2\mu}$. This characteristic is shared by other learning functions listed below. Second, our learning process is comparable to the learning functions used in the literature. Denote the *quality* of a machine as $q_i \equiv 1 - \theta_i$. In terms of q_i, (4) can be rewritten as

$$q_i' - q_i = (1 - e^{-\mu Y_i}) \cdot (1 - q_i).$$

When Y_i is exogenous and constant over time, $(1 - e^{-\mu Y_i})$ can be regarded as a constant parameter. This formulation corresponds to the learning function in Parente (1994). Arrow's (1962) specification amounts to

$$q_i = \alpha_1 \Lambda_i^{\alpha_2}, \quad \alpha_1, \alpha_2 > 0,$$

where Λ_i denotes the cumulative use of the machine. Jovanovic and Nyarko (1995) suggests the following learning process

$$q_i = 1 - \frac{1}{\varpi_1 + \varpi_2 \Lambda_i}, \quad \varpi_1, \varpi_2 > 0.$$

Our formulation can be rewritten as

$$q_i = 1 - \bar{\theta}_i \cdot \exp(-\mu \Lambda_i),$$

where $\bar{\theta}_i$ is the initial value of θ_i. Therefore, our formulation can be viewed as an alternative functional form for learning process.

With the current demand structure, the monopolist machine producer cannot sell the generation-i machine to more than $N[1 - F(m\theta_i)]$ users, and this limits the amount of quality improvement which comes from learning. Thus, the process of learning is based on the current value of θ_i, and it can be myopic. In contrast, R&D investment is driven by the future profit from demand expansion. In other words, R&D investment is based on the future *expectation*, while learning is based on the *history* of past quality improvement. The difference will be more evident in Section 4, where we investigate the implication of the skill distribution on the dynamics of the industry.

2.4 Equilibrium

Clearly, given the structure of demand, it is optimal for the monopolist to price $p_i = b_i$ so that he can capture all the surplus. With this pricing, he still faces the choice of how many machines of each generation to sell.

Bellman's Equation: Denote the monopolist's value function by $V(b_{i-1}, b_i, \theta_{i-1}, \theta_i)$. The state variables are the performance and the quality of the machines that he is currently producing. The Bellman's equation is

$$\begin{aligned} & V(b_{i-1}, b_i, \theta_{i-1}, \theta_i) \\ = \; & \max_{\theta'_{i-1}, \theta'_i, Y_i} \left\{ \Pi_i(b_i, \theta_i, \theta'_i, Y_i) + \Pi_{i-1}(b_{i-1}, \theta_{i-1}, \theta'_{i-1}, Y_i) + \beta W(b_{i-1}, b_i, \theta'_{i-1}, \theta'_i) \right\}, \end{aligned}$$

where

$$\begin{aligned} & W(b_{i-1}, b_i, \theta'_{i-1}, \theta'_i) \\ = \; & \max \left\langle \underbrace{V(b_{i-1}, b_i, \theta'_{i-1}, \theta'_i)}_{\text{do not innovate}}, \underbrace{\max_{b_{i+1}} \left\{ V(b_i, b_{i+1}, \theta'_i, \theta'_i + c(b_{i+1}/b_i)) \right\}}_{\text{innovate}} \right\rangle. \end{aligned}$$

The flow net profit of the monopolist from the ith-generation machines, $\Pi_i(b_i, \theta_i, \theta'_i, Y_i)$, is

$$\Pi_i(b_i, \theta_i, \theta'_i, Y_i) = \underbrace{b_i \cdot Y_i}_{\text{profit}} - \underbrace{b_i \cdot R(\theta_i e^{-\mu Y_i} - \theta'_i)}_{\text{R\&D cost}}.$$

The flow sales amount accruing from ith-generation machines, Y_i, is chosen optimally subject to

$$Y_i \in [0, N\{1 - F(m\theta_i)\}],$$

where $F(\cdot)$ is the cumulative distribution function of the skill distribution.

The flow net profit from the $(i-1)$th-generation machines, $\Pi_{i-1}(b_{i-1}, \theta_{i-1}, \theta'_{i-1}, Y_i)$, is

$$\Pi_{i-1}(b_{i-1}, \theta_{i-1}, \theta'_{i-1}, Y_i) = \underbrace{b_{i-1} \cdot Y_{i-1}}_{\text{profit}} - \underbrace{b_{i-1} \cdot R(\theta_{i-1} e^{-\mu Y_{i-1}} - \theta'_{i-1})}_{\text{R\&D cost}},$$

where $Y_{i-1} = \max\{0, N[1 - F(m\theta_{i-1})] - Y_i\}$. This transpires since the potential demand for the $(i-1)$th-generation machines is $1 - F(m\theta_{i-1})$ and it is always subject to the competition from the ith-generation machines.

Stationary Formulation: Let the degree of performance enhancement be represented by $\gamma_i \equiv b_i/b_{i-1}$. Then, from the homogeneity of the return functions Π_i and Π_{i-1}, it follows that $V(b_{i-1}, b_i, \theta_{i-1}, \theta_i) = b_{i-1} v(\gamma_i, \theta_{i-1}, \theta_i)$ and $W(b_{i-1}, b_i, \theta'_{i-1}, \theta'_i) = b_{i-1} w(\gamma_i, \theta'_{i-1}, \theta'_i)$. Let the subscripts h and l denote the higher- and the lower-technology machines that the firm is currently producing. The stationary formulation is:

$$v(\gamma, \theta_l, \theta_h) = \max_{\theta'_h, \theta'_l, Y_h} \left\{ P_h(\gamma, \theta_h, \theta'_h, Y_h) + P_l(\theta_l, \theta'_l, Y_h) + \beta w(\gamma, \theta'_l, \theta'_h) \right\}, \tag{P1}$$

$$w(\gamma, \theta'_l, \theta'_h) = \max \left\langle \underbrace{v(\gamma, \theta'_l, \theta'_h)}_{\text{do not innovate}}, \underbrace{\max_{\gamma'} \left\{ \gamma v(\gamma', \theta'_h, \theta'_h + c(\gamma')) \right\}}_{\text{innovate}} \right\rangle, \tag{5}$$

where

$$P_h(\gamma, \theta_h, \theta'_h, Y_h) = \underbrace{\gamma \cdot Y_h}_{\text{profit}} - \underbrace{\gamma \cdot R(\theta_h e^{-\mu Y_h} - \theta'_h)}_{\text{R\&D cost}},$$
$$Y_h \in [0, N\{1 - F(m\theta_h)\}],$$

and

$$P_l(\theta_l, \theta'_l, Y_h) = \underbrace{\max\{0, N[1 - F(m\theta_l)] - Y_h\}}_{\text{profit}} - \underbrace{R(\theta_l e^{-\mu \max\{0, N[1-F(m\theta_l)] - Y_h\}} - \theta'_l)}_{\text{R\&D cost}}.$$

To ensure the existence of the value functions, an additional constraint is imposed: $\gamma \leq \bar{\gamma}$, where $1 < \bar{\gamma} < \beta^{-1}$.

The following proposition obtains:

Proposition 2 *Suppose that $F(\cdot)$ is continuous. Then, a unique continuous function v which satisfies the Bellman's equation exists. v is weakly increasing in the first term, and weakly decreasing in the second and third terms.*

δ	σ	μ	m	N	ξ	β	κ	ζ
1	1	0.05	80	1	1000	0.8	0.018	3.5

Table 2: Parameter Values

Proof. See Appendix. ■

This result is intuitive: The value function is increasing in the relative performance of the higher technology and the quality levels of both technologies.

3 The Dynamics – Numerical Experiment

It is difficult to establish the properties of the model analytically. In this section, the dynamics of the model are analyzed through numerical simulations. For the skill distribution, a lognormal distribution is used:

$$\ln s \sim N(\delta, \sigma^2).$$

δ is the scale parameter and σ is the shape parameter of the distribution.

The R&D cost function is specified as follows:

$$R(\theta e^{-\mu Y} - \theta') = \xi \cdot (\theta e^{-\mu Y} - \theta')^2,$$

where $\xi > 0$ is a parameter. R&D cost is zero when $\theta' = \theta e^{-\mu Y}$; in that case only "learning by using" takes place. For the loss of quality with innovation, let

$$c(\gamma) = \kappa + \zeta(\gamma - 1)^2,$$

where $\kappa \geq 0$ and $\zeta > 0$ are parameters.

Table 2 shows the values of the parameter used for the experiment. Standard value function iteration method is used for the computation. A simple grid search method is utilized (since the problem is not necessarily concave, a method which involves differentiation is hard to use) and linear interpolation is employed when necessary.[17] Initial values are given as $\theta_1 = 0.01$, $\theta_2 = 0.05$, and $\gamma = 1.05$.

It turns out that in the experiments below, it is optimal for the monopolist to set $Y_h = 1 - F(m\theta_h)$, that is, to sell the higher-technology machine as many users as possible at any time. This is

[17] Interpolation is needed to compute the value along the "innovate" branch. Better interpolation methods, e.g. cubic interpolations, do not alter the result.

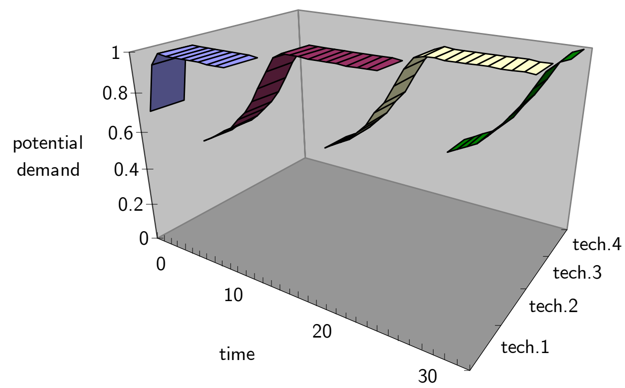


Figure 3: Diffusion Curves

intuitive for three reasons. First, since $\partial P_h / \partial Y_h = \gamma$ and $\partial P_l / \partial Y_l = 1$, higher-technology machines give higher (static) marginal profit to the producer. Second, since $\partial \left(\theta e^{-\mu Y}\right) / \partial Y = -\mu \theta e^{-\mu Y}$, keeping Y constant, the contribution of Y to learning is larger when θ is larger. Since it is always the case that $\theta_h \geq \theta_l$, learning efficiency is higher in the production line of the higher-technology machine.[18] Third, the value function when innovation occurs, $\max_{\gamma'} \left\{\gamma v(\gamma', \theta_h', \theta_h' + c(\gamma'))\right\}$, depends on θ_h', but not θ_l'. This gives an additional incentive to make θ_h smaller.

The next two sections show the dynamics of technology diffusion and adoption in the model. First, we focus on the process of diffusion itself. Second, adoption timings of the users with different skills are described.

3.1 Time Series for Technology Diffusion and Innovation

Figure 3 shows the time series for technology diffusion. The four curves represent the potential demand $D_i(p_i, \theta_i)$ for the four technologies. The price p_i is constant at b_i (see Section 2.4), but

[18] It is also true that $\mu \theta e^{-\mu Y}$ is decreasing in Y, so this intuition goes the opposite direction when $Y_h \geq Y_l$. However, $\mu \theta e^{-\mu Y}$ is bounded below by $\mu \theta$ for the change in Y, and thus learning efficiency is higher in high-tech if $\mu \theta_h e^{-\mu Y_h} \geq \mu \theta_l$ holds. This appears to hold almost always, since θ_l is always very close to zero.

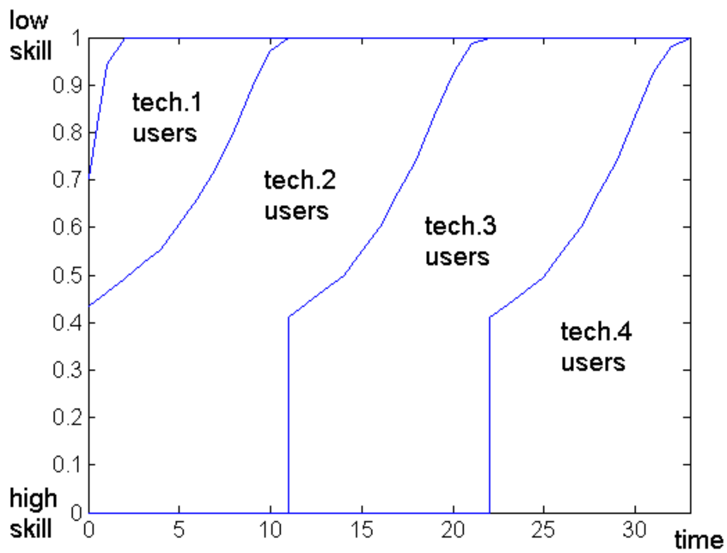


Figure 4: Diffusion Curves: Another Look

$D_i(p_i, \theta_i)$ increases over time since θ_i declines. The diffusion curves typically exhibit an S-shape, which is observed in many empirical literature of technology diffusion.[19] Figure 4 provides another look at the diffusion curves. It shows what fraction of the users use each technology in each period. Since high-skilled users adopt a new technology first, we can consider that the highest-skilled user is at the bottom and the lowest-skilled user is at the top of the figure. By reading the figure horizontally, we can see how a user at each percentile of skill distribution changes his technology over time. Figure 5 shows the dynamics of the error ratio, θ. They fall almost at a constant speed. The S-diffusion curves then follow from the shape of the cumulative distribution function for the lognormal skill distribution.

An innovation occurs when θ_h becomes nearly zero, that is, when the best-practice machine has diffused to nearly everyone. This transpires since the cost of innovation is lower when θ_h is small. The cost of bringing the $(i+1)$th generation technology on line is twofold. First, the lower-technology machine switches from the $(i-1)$th generation to ith generation. This implies that the quality of the lower technology changes from θ_{i-1} to θ_i. Usually this entails a deterioration in quality, since most of the time $\theta_{i-1} \leq \theta_i$ holds. If θ_i (which is θ_h before the innovation) is nearly

[19] For a recent survey of the models which exhibit S-diffusion curve, see Geroski (2000).

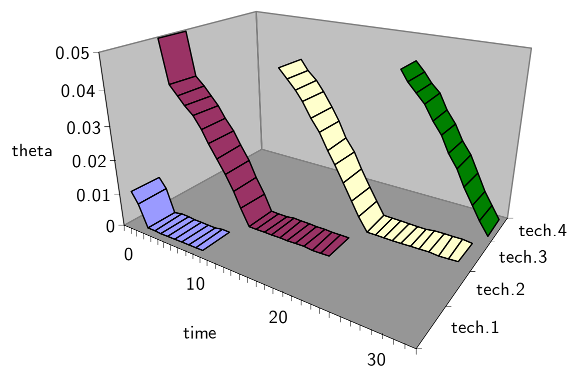


Figure 5: Error Ratio

zero, this cost is also very small. Second, the error ratio of the higher-technology machine increases by $c(\gamma)$. That is, $\theta_{i+1} = \theta_i + c(\gamma)$. This deterioration in quality entails a loss of demand. The cost of this change (the loss of demand) depends on how many users are in between the skill levels $m\theta_i$ and $m[\theta_i + c(\gamma)]$. Since our skill distribution is lognormal, $f(m\theta_i)$ is increasing in θ_i when θ_i is small enough. Consequently, the loss of demand, $\int_{m\theta_i}^{m[\theta_i+c(\gamma)]} f(x)dx$, is small when θ_i is small.

The fact that innovation occurs when $\theta_h \approx 0$ implies that, if θ_h approaches to zero quickly (which results in faster diffusion), the speed of innovation is also fast. There is a connection between diffusion and the *timing* of innovation.

3.2 Heterogeneity in Technology Adoption

Figure 6 shows the technology level of machines that the users of each skill level have adopted.[20] The skill levels are s =1, 2, 3, and 4 in this example. Higher-skilled users adopt the new technology earlier. In period 5, for example, the users with skill levels 3 and 4 use the second technology, while the users with skill levels 1 and 2 use the first technology, which has less performance. In period 6, the users with skill level 2 adopt the second technology as its quality improves, while the users

[20] The minimum is set to 1 in the figure to clarify the comparison.

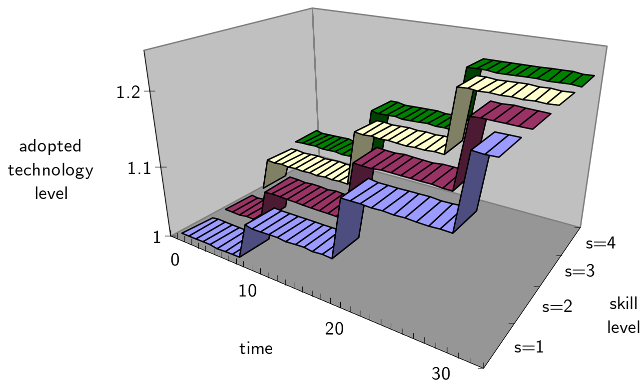


Figure 6: Technology Adoption

with skill level 1 keeps using the first technology.

The difference in adoption timing and the difference in technology level are both determined by the machine producer's decision about θ and γ. The values of θ and γ are determined endogenously, and they are affected by the distribution of skills. Through this channel, the other people's skill levels affect a user's own technology level.

One of the criticisms that Jovanovic (1998) made on vintage capital models, such as Arrow (1962b) and Parente (1994), is that they exhibit unrealistic leapfrogging: in these models, a "productivity miracle" and a "productivity disaster" repeat endlessly. The data do not show this type of leapfrogging at any level of aggregation. Our model passes Jovanovic's critique; there is no turnover in productivity distribution among users. To see this, let's look at Figure 6 again. The productivity of each user is represented by the height of the graph (the level of adopted technology). A low-skilled user adopts a new technology some time after a high-skilled user's adoption, and a "productivity catch-up" occurs. However, the low-skilled user's productivity never exceeds the high-skilled user's productivity (therefore, a leapfrogging never happens), since a high-skilled user always adopts a high-performance technology earlier.

4 The Effect of Skill Distribution

In this section, the impact of the shape of skill distribution on the dynamics is analyzed. An implication of the model is that the *shape* of the skill distribution can have a nontrivial effect on the dynamics of technology development and diffusion. A user's dynamic path of performance is not only governed by his own skill level, but also by the economywide distribution of skills.

4.1 Simple Analytical Examples

First, the basic mechanism is illustrated by a series of simple examples. Only one important element is contained in each model so that analytical procedures can be employed.

4.1.1 Learning and Diffusion

The first model analyzes the process of learning. Assume that there exists only one technology: call it technology 1. Technology 1 arrives at time 0. The next generation technology never arrives. Let the density function $f(x) \equiv F'(x)$ for the skill distribution be uniform:

$$f(x) = \begin{cases} 1/\sigma & \text{for } x \in [\delta - \sigma/2, \delta + \sigma/2]\,, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $\delta \in (0, 1)$ and $\sigma \in (0, 2\delta)$. The mean of the distribution is δ, and the variance is $\sigma^2/12$. δ shifts the distribution in a parallel fashion, while σ determines the shape of the distribution. Denote the error ratio of technology 1 by θ and its performance by b. Technology 1 is owned by the capital goods producer, who learns from the users. The quality improvement of technology 1 (decline of θ) follows

$$\theta' = \theta \cdot e^{-\mu Y}, \tag{7}$$

where Y is the total use of technology 1. The optimal policy for the monopolist is to sell $Y = N[1 - F(m\theta)]$ with the price b. Thus (7) can be rewritten as

$$\theta' = \theta \cdot \exp\{-\mu \cdot N[1 - F(m\theta)]\}. \tag{8}$$

Suppose that technology 1 starts from $\bar{\theta}$. It is assumed that $m\bar{\theta} < \delta + \sigma/2$ (otherwise, the quality improvement does not start). The following proposition characterizes the relation between dispersion parameter σ and the speed of quality improvement.

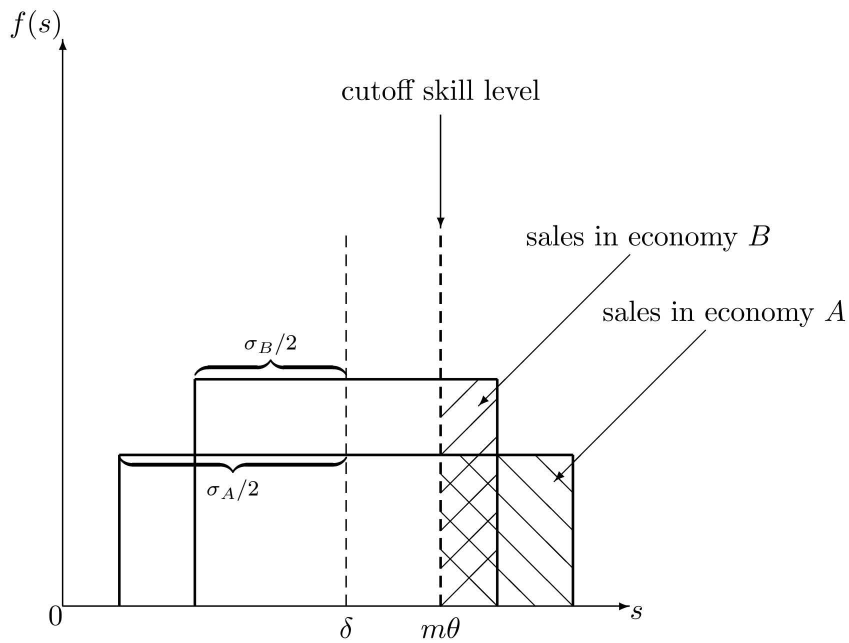


Figure 7:

Proposition 3 *Consider two economies A and B with skill distributions F_A and F_B. F_A and F_B have a common δ and two different σ's, σ_A and σ_B, where $\sigma_A > \sigma_B$. For given θ, the speed of quality improvement, θ/θ', is larger in economy A if and only if $\theta > \delta/m$.*

Proof. See Appendix. ■

Thus, a more "unequal" economy A experiences a relatively faster diffusion initially (when θ is large), while economy B experiences a faster diffusion later on (when θ is small). Figure 7 shows the equilibrium graphically. When the error ratio is θ, the sales of each economy is proportional to the shaded area. The area is larger for A if and only if the cutoff skill level $m\theta$ is larger than δ. Since the amount of learning is a increasing function of the sales, the proposition follows.[21]

[21] From (8), the following result is also straightforward: if the skill distribution of economy X first-order stochastically dominates the skill distribution of economy Y, the speed of quality improvement is faster in economy X for given θ.

4.1.2 R&D and Diffusion

Here it is assumed that the machine producer can improve quality by R&D investment. Again, we consider the diffusion of a single technology, with the same notation as the previous section. The skill distribution is given by (6). Machine production is monopolized perpetually. There is no learning, but the monopolist can improve the quality of machines by R&D investment. He chooses the new quality $(1-\theta')$ of the technology (where $\theta' \leq \theta$) subject to the R&D cost function

$$R(\theta-\theta') = \xi \cdot (\theta-\theta')^2.$$

Clearly the optimal sales policy for the monopolist is to sell $N[1-F(m\theta)]$ with the price b, which yields the profit of $N[1-F(m\theta)] \cdot b$. Then the only decision for him is the choice of θ'. Since it is not beneficial to set $\theta < \underline{\theta} \equiv \delta - \sigma/2$, it is innocuous to impose a constraint $\theta' \geq \underline{\theta}$. Suppose that the initial value of $\bar{\theta}$ satisfies $\bar{\theta} \leq \delta + \sigma/2$.

The monopolist's Bellman's equation is

$$v(\theta) = \max_{\underline{\theta} \leq \theta' \leq \theta} \{N[1-F(m\theta)] \cdot b - \xi(\theta-\theta')^2 + \beta v(\theta')\}.$$

The Euler equation is

$$\underbrace{\beta\left[\frac{Nbm}{\sigma} + 2\xi(\theta'-\theta'')\right]}_{\text{marginal benefit of reducing } \theta'} - \underbrace{2\xi(\theta-\theta')}_{\text{marginal cost}} - \lambda = 0, \tag{9}$$

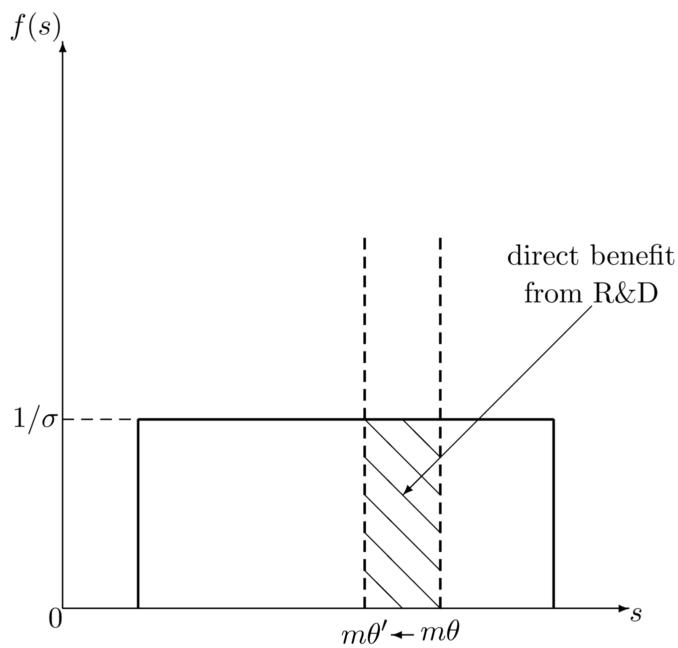
where λ is the Lagrange multiplier associated with the constraint $\theta' \geq \underline{\theta}$. With the initial condition $\bar{\theta}$ and the transversality condition $\lim_{t\to\infty} \beta^t \lambda_t = 0$, the unique solution of this second-order difference equation is (θ_t refers to the value of θ at period t):

$$\theta_t = \begin{cases} \bar{\theta} - \phi t & \text{if } t \leq (\bar{\theta} - \underline{\theta})/\phi, \\ \underline{\theta} & \text{otherwise,} \end{cases}$$

where

$$\phi \equiv \frac{\beta N b m}{2(1-\beta)\sigma\xi}$$

governs the speed of quality improvement. The speed of quality improvement is higher when b, m, N, and $1/\sigma$ are larger, since these raise the marginal benefit from quality improvement. The parameter β affects since the benefit of R&D activity is an "investment". A rise in the R&D cost



$f(s)$
direct benefit
from R&D
$1/\sigma$
0
s
$m\theta' \leftarrow m\theta$

Figure 8:

parameter, ξ, slows the diffusion. A notable result is that the speed of quality improvement, ϕ, is decreasing in σ. In an "unequal" economy, the speed of quality improvement due to R&D is slower.

Figure 8 shows the equilibrium. The area of shaded region is proportional to the direct benefit from R&D, which is the first term in the square bracket in (9). This area represents the amount of (additional) demand that can be captured by the quality improvement. The area is larger if the density of the skill distribution is "thick": in the case of uniform distribution, the area is proportional to $1/\sigma$. Thus, the incentive for quality improvement is larger when σ is smaller.

4.1.3 Size of Innovation

The third model focuses on the decision to innovate. Assume that the skill distribution is given by (6), but $\delta - \sigma/2 \leq 0$. Here, quality improvement is exogenous: $\theta_t = \rho^t \cdot \bar{\theta}$, where $\rho \in (0,1)$. Let $m\bar{\theta} \leq \delta + \sigma/2$ so that the demand for the machines is

$$D_t = N \cdot \left(\delta + \frac{\sigma}{2} - m\theta_t\right) \cdot \frac{1}{\sigma}.$$

The monopolist can choose the level of the machine performance, γ, when he innovates. It is assumed that if he chooses a higher γ, then initial quality is lower. Thus,

$$\bar{\theta} = c(\gamma), \text{ where } c'(\cdot) > 0.$$

This imposes a trade-off: if the monopolist chooses a revolutionarily high-performance technology, he can sell it at a high price (price equals γ) but demand is going to be low because of the low quality. A low-performance technology would sell to many, but the profit per machine is small.

The monopolist's problem is

$$\max_{\gamma} \quad \sum_{t=0}^{\infty} \beta^t \cdot \gamma \cdot N \cdot \left[\delta + \frac{\sigma}{2} - m \cdot \rho^t \cdot c(\gamma)\right] \cdot \frac{1}{\sigma}.$$

The objective function can be arranged to reduce this problem to:

$$\max_{\gamma} \quad \frac{\gamma N}{\sigma} \left[\frac{1}{1-\beta}\left(\delta + \frac{\sigma}{2}\right) - \frac{m \cdot c(\gamma)}{1-\beta\rho}\right].$$

The first order condition is

$$\underbrace{\frac{N}{\sigma}\left[\frac{1}{1-\beta}\left(\delta + \frac{\sigma}{2}\right) - \frac{m \cdot c(\gamma)}{1-\beta\rho}\right]}_{\text{marginal benefit from increasing } \gamma} - \underbrace{\frac{\gamma N}{\sigma}\frac{m \cdot c'(\gamma)}{1-\beta\rho}}_{\text{marginal cost}} = 0. \tag{10}$$

The next proposition follows.[22]

Proposition 4 *Optimal performance γ is increasing in σ.*

Proof. See Appendix. ■

Since γ is increasing in σ, $\bar{\theta}$ is also increasing in σ. In an "unequal" economy, innovation is more drastic, and the initial quality of machines is low when introduced.

4.2 Numerical Simulation

Next, the full model is analyzed numerically. The main focus is the effect of the change of the shape parameter σ.

4.2.1 Diffusion and the Distribution of Skills

Figure 9 compares the diffusion of a technology with different values for σ. The other parameter values are the same as in Table 1. Three cases are displayed: $\sigma = 0.8$, $\sigma = 1.0$, and $\sigma = 1.2$. It can be seen that the diffusion is faster with a small σ. The diffusion curves start from almost the same level of adoption (about 40%), but it takes 5 periods when $\sigma = 0.8$, 6 periods when $\sigma = 1.0$, and almost 8 periods when $\sigma = 1.2$ to diffuse to 70% of users.

Two effects are working. First, as we saw in Section 4.1.2, the incentive for R&D is larger when the skill distribution is less dispersed. R&D investment in first 6 periods[23] is 0.085 for $\sigma = 0.8$, 0.081 for $\sigma = 1.0$, and 0.068 for $\sigma = 1.2$. This leads to faster diffusion for smaller σ. Second, as in Section 4.1.1, learning is faster in the beginning the larger σ is, but it favors the smaller σ later. This second effect can be seen in the figure from the fact that in the first few periods, the speed of diffusion is almost the same for all the distributions. The combination of these two yields the

[22] For a general skill distribution, the first order condition becomes

$$N \sum_{t=0}^{\infty} \beta^t f[m \cdot \rho^t \cdot c(\gamma)] \left\{ \frac{1}{H[m \cdot \rho^t \cdot c(\gamma)]} - m \cdot \rho^t \cdot c'(\gamma) \cdot \gamma \right\} = 0,$$

where $H(x) = f(x)/[1 - F(x)]$ is the hazard function. Thus, the optimal choice of γ is closely related to the property of $H(x)$. An important fact here is that $H(x)$ is decreasing in σ for the uniform distribution employed in this section. This property holds for the lognormal distribution used in the numerical experiments, when x is not too small (see Hastings 1975, p.87).

[23] The R&D investment in later periods do not contribute much to the diffusion of current technology. (It is rather for the next upcoming technology, since the next technology starts from the quality level $\theta_h + c(\gamma)$.) So we do not take it into account here.

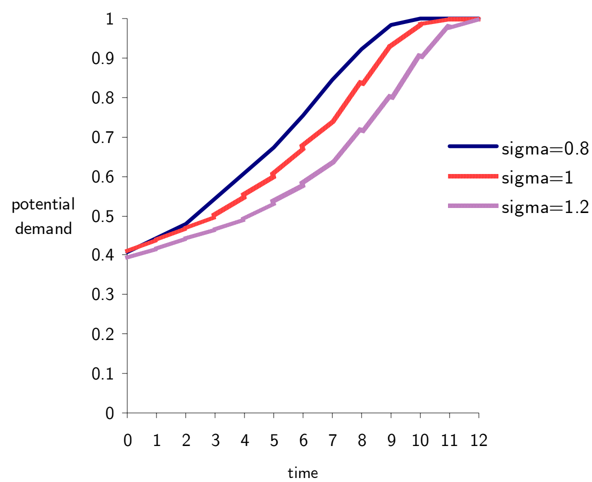


Figure 9: Diffusion Curves

overall faster diffusion for smaller σ. This is consistent with the empirical works cited in Oi (1997): a homogenous society exhibits faster diffusion.

4.2.2 Innovation and Growth

The innovation decision is also affected by the shape parameter σ. There are two decisions to make: when to innovate and how much to innovate. Figures 10 and 11 show that there is a systematic relationship between σ and the innovation decision. As σ increases, the degree of performance enhancement, γ, becomes larger. This is consistent with the analytical result in Section 4.1.3. It reflects the fact that when σ is large, the marginal loss from larger θ (low quality) is smaller (relative to the marginal gain). The periods between innovations become longer as σ gets larger. As is seen in Section 3.1, innovation occurs when $\theta_h \approx 0$. In an economy with large σ, it takes time for θ_h to become near zero.[24]

In a stationary equilibrium, the rate of growth in machine performance for each user is dictated by γ and the interval between innovation (denote it by T). In fact, the average rate of growth is

[24] This is due to two effects. First, a larger θ_h (due to a larger γ) in each innovation implies that it takes time to make θ_h small. Second, as is shown in Section 4.2.1, the diffusion is slower when σ is larger.

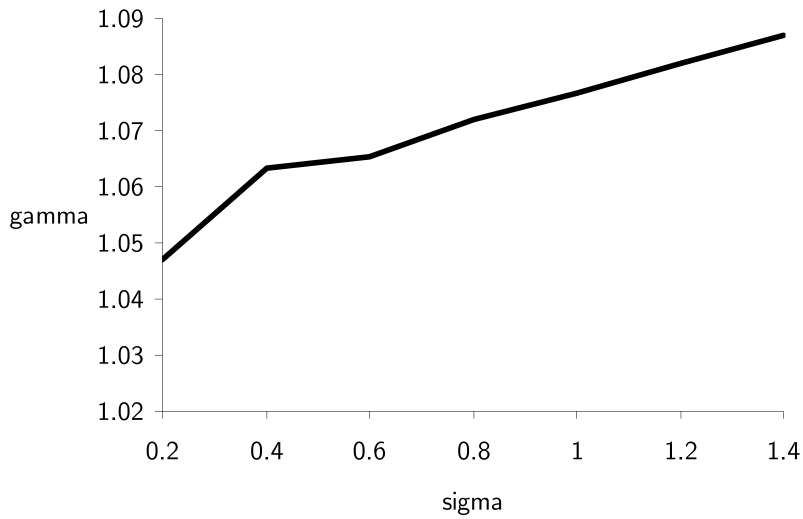


Figure 10: Size of Innovations

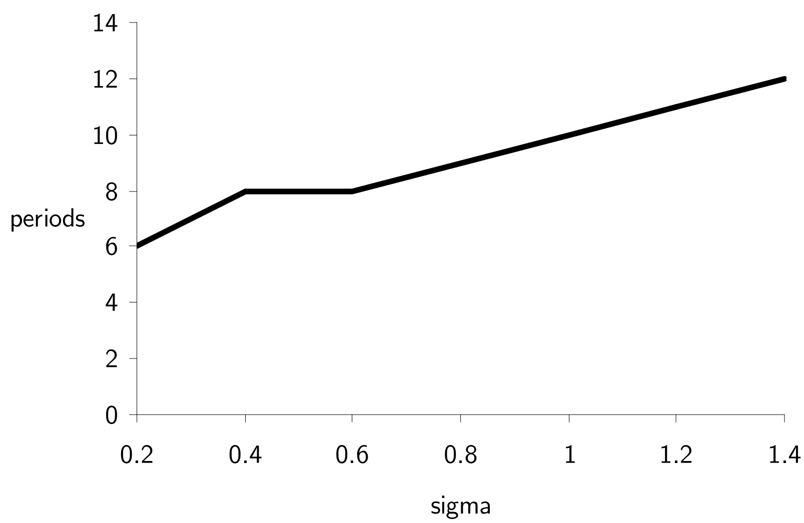


Figure 11: Periods between Innovations

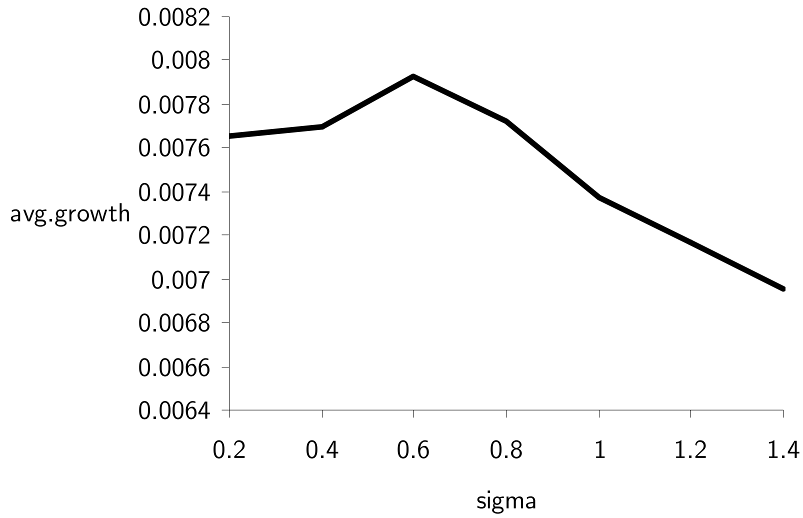


Figure 12: Average Growth Rate

common among the active users and given by $g = (\log \gamma)/T$. Figure 12 plots the change in average growth rate g as σ changes from 0.4 to 1.4. There is a non-monotonicity, as the effects of γ and T offset each other. When σ is very low, the innovations are too incremental. On the other hand, when σ is very high, even though each innovations are revolutionary, the time intervals between innovations are too long.

5 Other Implications

5.1 General Equilibrium

Our analysis has been in a partial equilibrium framework. Extending it to a general equilibrium model would complicate the analysis, but it maintains the interesting features.

To illustrate, consider a simpler model where there are only two skill levels. The skilled workers (H) can engage in R&D (for quality improvement) in the capital goods sector (H^R) or work as machine users (H^U) in the consumption goods sector. All the unskilled workers (L) become machine users (L^U). The resource constraint is

$$H^R + H^U = H, \tag{11}$$

and

$$L^U = L. \tag{12}$$

Suppose that the supply of unskilled labor (L) increased suddenly. From (12) L^U increases, so keeping H^U constant, the relative amount of unskilled users of machine increases. Thus it would be more beneficial for capital good producers to improve the quality of machines. This increases H^R, and from (11), it reduces H^U. Then, the relative amount of unskilled users increases further. The general equilibrium effect makes quality improvement even faster.

Another interesting implication would arise when we consider multiple industries. Industries may vary in their timing of innovation. Since skill is rewarded the most in those industries which have just experienced innovation, skilled workers tend to be allocated to those industries and would move out of them as the technologies mature. It adds another dynamic aspect to the model. If the size of innovation differs among the industries, the most skilled workers would move to the industry whose innovation is the most revolutionary. Thus, the skill distribution of potential users in an industry is affected by the behavior of the other industries.

5.2 Growth and World Income Distribution

Our model applied to the world economy can shed a light on world growth and the evolution of world income distribution. As is described in Section 3.2, there can be a large difference in the timing of new technology adoption due to differences in skills. It is well known that there is large difference in the educational attainment between the advanced countries (the North) and the less developed countries (the LDCs). Even if the LDCs have an access to new technologies by importing new machines, they may not adopt them because of the lack of skills. LDCs adopt new technologies only after they mature and become easier to be handled. The difference in skills can be a source of heterogenous technology adoption, and in turn, large TFP differences.[25]

[25] Acemoglu and Zilibotti (2001) also attempts to explain the differences in TFP in the world where LDCs have an access to the same set of technologies as the North. They argue that the technologies tend to be inappropriate for the LDCs, since they are targeted at the use in the skill-abundant North and tend to be complement to skills. In our model, the nature of technology (skill-complementarity) changes over time. In the Acemoglu and Zilibotti model, a machine that is skill-complementary can never be successfully adopted by unskilled workers, while in our model an adoption is possible if a sufficient amount of quality improvement occurs.

One interesting aspect of our model is that, even if the Northern countries are the only ones who produce new capital goods and conduct R&D, technological progress and the world growth rate depend on the skill distribution of the whole world. As we have seen above, diffusion and innovation are interrelated, and the rate of diffusion affects the incentive for innovation. It is possible that increases in educational expenditure in the LDCs not only speed up the rate of diffusion but also drive up the world growth rate.

5.3 Skill Premium

In our model, the demand curve (2) is horizontal. It follows that the monopolist captures all the surplus and no surplus is left to the users. Then, all users are left with zero surplus, regardless of their skill levels. This is the result of our simple specification. If $g_i(s, \theta_i)$ in (1) is specified differently, some users will be left with some surplus. For example, let

$$g_i(s, \theta_i) = \psi \frac{\theta_i}{s},$$

where ψ is a constant parameter. Then a user with skill level s buys the i-generation machine if $b_i - \psi\theta_i/s - p_i \geq 0$, that is, $s \geq \psi\theta_i/(b_i - p_i)$. The potential demand becomes

$$D_i(p_i, \theta_i) = N\left[1 - F\left(\frac{\psi\theta_i}{b_i - p_i}\right)\right].$$

This demand curve is downward-sloping. Suppose that the equilibrium price is p_i^*. If $s > \psi\theta_i/(b_i - p_i^*)$, then $b_i - \psi\theta_i/s - p_i^* > 0$ and the user obtains a positive surplus from using the machine. Moreover, the surplus is increasing in s. A user with higher skill level enjoys a larger amount of surplus. We can interpret this difference in the surplus as *skill premium*.

Skill premium changes over time. It changes for the users who operate generation-i machines, as θ_i and p_i changes. The decision of switching to a new technology also affects the skill premium. The full analysis of the modified model is substantially complicated, and it is beyond the scope of this paper.

6 Conclusion

This paper has presented a theory of technology diffusion. We started from two assumptions. First, high level of skills are required to adopt a new technology. Second, the nature of a new technology changes over time, so that it becomes more and more accessible to less skilled machine users. This change is called a quality improvement, and it is carried out by the learning and R&D investment of the capital good producer.

We considered a specific form of learning: learning by using. A learning function is developed in Lemma 1, and applied to our model. The dynamics of diffusion show an S-shape. Since the speed of quality improvement depends on the skill distribution of the economy, the timing of adoption by each user is also affected by the distribution of skills. In Section 4, we analyzed the relation between the skill distribution and diffusion.

The machine producer is also allowed to start producing a new generation of machines. It is called innovation. If he chooses to produce high-performance machines, he has to start from low quality. Thus the timing and the size of innovation is also influenced by the skill distribution. In our numerical experiment, larger dispersion of skill delays innovation, and makes the size of innovation larger.

This framework is capable of addressing many issues. Some directions were discussed in Section 5. Another interesting extension is to endogenize skill formation. In this paper, the skill distribution is treated as exogenous. In reality, there is a substantial portion of skill that is the result of intentional investment.[26] On the one hand, as discussed in Section 5.3, the return to skill depends on the dynamics of innovation and diffusion. On the other hand, in Section 4, it has been shown that skill distribution affects innovation and diffusion. With endogenous skill formation, the three factors – diffusion, innovation, and skill investment – interact with each other.[27] Detailed investigation is left to future research.

[26] Schultz (1975) emphasizes the role of education in enhancing the ability of students to perceive new classes of problems, to clarify such problems, and to learn ways of solving them. See also Welch (1970).

[27] Redding (1996) constructs a simple endogenous growth model with R&D and human capital investment. New technologies (as the results of R&D) and human capital are complementary, and he shows that, when there is an indivisibility in R&D, multiple equilibria can exist.

Appendix

A Proofs

Proof of Lemma 1:[28]

Denote the number of the errors not found before nth user by J_n. J_n is a random variable. By definition, $J_1 = J$. Let the (unconditional) expected probability that nth user finds an error be π_n. Clearly $\pi_1 = J/K$. It is also clear that $\pi_n = E[J_n]/K$, where $E[\cdot]$ is an expectation operator. Let the event $A_n = \{n\text{th user finds an error}\}$ and $A_n^c = \{n\text{th user does not find an error}\}$. The following holds from the law of iterated expectations:

$$\begin{aligned}
E[J_n] &= E[J_n|A_{n-1}] \cdot \Pr[A_{n-1}] + E[J_n|A_{n-1}^c] \cdot \Pr[A_{n-1}^c] \\
&= E[J_n|A_{n-1}] \cdot \pi_{n-1} + E[J_n|A_{n-1}^c] \cdot (1 - \pi_{n-1}) \\
&= (E[J_n|A_{n-1}] - E[J_n|A_{n-1}^c])\pi_{n-1} + E[J_n|A_{n-1}^c].
\end{aligned}$$

Clearly the first term is $-\pi_{n-1}$ (since one error is found if A_{n-1} happens and no error is found if A_{n-1}^c happens), and the second term is $E[J_{n-1}]$ (since nothing changes from the past user if the error is not found). Then, noting $\pi_{n-1} = E[J_{n-1}]/K$,

$$\begin{aligned}
E[J_n] &= E[J_{n-1}]\left(1 - \frac{1}{K}\right) \\
&= J\left(1 - \frac{1}{K}\right)^{n-1}. \qquad (13)
\end{aligned}$$

Consider the situation where J, K, and Z are large. Especially, keep $\theta = J/K$ and $z = Z/K$ constant and let $J, K, Z \to \infty$. Then at the limit, the expected ratio of remaining errors goes to

$$\begin{aligned}
\lim \frac{E[J_{zK}]}{K} &= \lim \frac{J}{K}\left(1 - \frac{1}{K}\right)^{zK-1} \\
&= \theta \cdot e^{-z}.
\end{aligned}$$

■

Proof of Proposition 2:

Let the right-hand-side of (P1), combined with (5), define a mapping T. Note that P_h and P_l are

[28] I thank Shakeeb Khan for helping the development of this lemma.

bounded and continuous, and the constraint correspondence for θ'_h, θ'_l, Y_h, and γ' is compact-valued and continuous. Then from the standard argument, T defines a contraction mapping on a space of bounded and continuous functions and the existence, uniqueness, and continuity of the fixed point v follows (Stokey and Lucas with Prescott 1989, henceforth SLP, Theorem 4.6). The discounting part of Blackwell's sufficient condition (SLP Theorem 3.3) is ensured by the assumption $\beta\bar{\gamma} < 1$.

For the increasingness in γ, first note that since P_h and P_l are nonnegative, v is also nonnegative. Then, if v is increasing in γ, Tv is also increasing in γ, since the domain of the choice variables are not affected by γ. From the Corollary 1 of Theorem 3.2 in SLP, the fixed point of T is also increasing in γ.

For the decreasingness in θ_h and θ_l, pick $(\theta_{h1}, \theta_{l1})$ and $(\theta_{h2}, \theta_{l2})$ where $\theta_{h1} \geq \theta_{h2}$ and $\theta_{l1} \geq \theta_{l2}$. For each γ, let the optimal choice of $(\theta'_h, \theta'_l, Y_h, \gamma')$ under $(\theta_{h1}, \theta_{l1}, \gamma)$ be $(\theta'^*_h, \theta'^*_l, Y^*_h, \gamma'^*)$. Assume that $v(\gamma, \theta_h, \theta_l)$ is decreasing in θ_h and θ_l. Then $w(\gamma, \theta'_h, \theta'_l)$ is decreasing in θ'_h and θ'_l. Let $\tilde{\theta}'_h \equiv \min\{\theta_{h2} e^{-\mu Y^*_h}, \theta'^*_h\}$ and $\tilde{\theta}'_l \equiv \min\{\theta_{l2} e^{-\mu \max\{0, 1-F(m\theta_{l2})-Y^*_h\}}, \theta'^*_l\}$. Then

$$
\begin{aligned}
& P_h(\gamma, \theta_{h2}, \tilde{\theta}'_h, Y^*_h) + P_l(\theta_{l2}, \tilde{\theta}'_l, Y^*_h) + \beta w(\gamma, \tilde{\theta}'_l, \tilde{\theta}'_h) \\
\geq\ & P_h(\gamma, \theta_{h1}, \theta'^*_h, Y^*_h) + P_l(\theta_{l1}, \theta'^*_l, Y^*_h) + \beta w(\gamma, \theta'^*_l, \theta'^*_h)
\end{aligned}
$$

follows, and $(\tilde{\theta}'_h, \tilde{\theta}'_l, Y^*_h, \gamma'^*)$ is feasible under $(\theta_{h2}, \theta_{l2}, \gamma)$. Thus $Tv(\gamma, \theta_{h2}, \theta_{l2}) \geq Tv(\gamma, \theta_{h1}, \theta_{l1})$. From the Corollary 1 of Theorem 3.2 in SLP, the fixed point of T is also decreasing in θ_h and θ_l. ■

Proof of Proposition 3:

From (4), θ/θ' is governed by Y_i, which, in this case, is $[(\delta + \sigma/2) - m\theta]/\sigma$. It is straightforward to check that $[(\delta + \sigma_A/2) - m\theta]/\sigma_A > [(\delta + \sigma_B/2) - m\theta]/\sigma_B$ if and only if $\theta > \delta/m$. ■

Proof of Proposition 4:

Define

$$
\Omega(\sigma) \equiv \frac{1}{1-\beta}\left(\delta + \frac{\sigma}{2}\right), \tag{14}
$$

and

$$
\Delta(\gamma) \equiv \frac{m \cdot c(\gamma)}{1-\beta\rho} + \gamma \frac{m \cdot c'(\gamma)}{1-\beta\rho}.
$$

Then the first order condition (10) can be rewritten as

$$
\Omega(\sigma) - \Delta(\gamma) = 0.
$$

From the implicit function theorem,

$$\frac{d\gamma}{d\sigma} = \frac{\Omega'(\sigma)}{\Delta'(\gamma)}.$$

From (14), $\Omega'(\sigma) > 0$. From the second order condition, $\Delta'(\gamma)$ is positive. Thus

$$\frac{d\gamma}{d\sigma} > 0.$$

■

References

[1] Abramovitz, M., "Resource and Output Trends in the United States Since 1870," *American Economic Review* 46 (1956), 5–23.

[2] Acemoglu, D., "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature* 40 (2002), 7–72.

[3] Acemoglu, D. and F. Zilibotti, "Productivity Differences," *Quarterly Journal of Economics* 116 (2001), 563–606.

[4] Anderson, H. D. and P. E. Davidson, *Occupational Trends in the United States* (Stanford, CA: Stanford University Press, 1940).

[5] Arrow, K. J., "The Economic Implications of Learning by Doing," *Review of Economic Studies* 29 (1962), 155–173.

[6] Barro, R. J. and X. Sala-i-Martin, *Economic Growth* (New York: McGraw-Hill, 1995).

[7] Bartel, A. P. and F. R. Lichtenberg, "The Comparative Advantage of Educated Workers in Implementing New Technology," *Review of Economics and Statistics* 69 (1987), 1–11.

[8] Cahners Electronics Group, *Electronics Industry Year Book 2000 Edition* (2000).

[9] Caselli, F. and W. J. Coleman II, "Cross-Country Technology Diffusion: The Case of Computers," *American Economic Review* 91 (2001), 328–35.

[10] Chari, V.V. and H. Hopenhayn, "Vintage Human Capital, Growth, and the Diffusion of New Technology," *Journal of Political Economy* 99 (1991), 1142–65.

[11] Denison, E. F., *Trends in American Economic Growth, 1929–1982* (Washington, D.C.: Brookings Institution, 1985).

[12] Doms, M., T. Dunne, and K. R. Troske, "Workers, Wages, and Technology," *Quarterly Journal of Economics* 112 (1997), 253–290.

[13] Galor, O. and D. Tsiddon, "Technological Progress, Mobility, and Economic Growth," *American Economic Review* 87 (1997), 363–82.

[14] Geroski, P. A., "Models of Technology Diffusion," *Research Policy* 29 (2000), 603–625.

[15] Goldin, C. and L. F. Katz, "The Origins of Technology-Skill Complementarity," *Quarterly Journal of Economics* 113 (1998), 693–732.

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

[17] Hastings, N. A. J., *Statistical Distributions: a Handbook for Students and Practitioners* (London: Butterworth, 1975).

[18] Jovanovic, B., "Vintage Capital and Inequality," *Review of Economic Dynamics* 1 (1998), 497–530.

[19] Jovanovic, B. and S. Lach, "Entry, Exit, and Diffusion with Learning by Doing," *American Economic Review* 79 (1989), 690–99.

[20] Jovanovic, B. and G. MacDonald, "Competitive Diffusion," *Journal of Political Economy* 102 (1994), 24–52.

[21] Jovanovic, B. and Y. Nyarko, "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves," *Brookings Papers: Microeconomics* (1995), 247–305.

[22] Klenow, P. and A. Rodríguez-Clare, "The Neoclassical Revival in Growth Economics: Has It Gone Too Far?," in B. S. Bernanke and J. J. Rotenberg, eds., *NBER Macroeconomics Annual 1997* (Cambridge: MIT Press, 1997).

[23] MacLeod, C., "Strategies for Innovation: the Diffusion of New Technology in Nineteenth-Century British Industry," *Economic History Review* 65 (1992), 285–307.

[24] Mansfield, E., *The Economics of Technological Change* (New York: W. W. Norton, 1968).

[25] Nelson, R. R., M. Peck, and E. Kalachek, *Technology, Economic Growth, and Public Policy* (Washington D.C.: Brookings, 1967).

[26] Nelson, R. R., and E. S. Phelps, "Investment in Humans, Technological Diffusion, and Economic Growth," *American Economic Review* 56 (1966), 69–75.

[27] Oi, W. Y., "The Welfare Implications of Invention," in T. F. Bresnahan and R. J. Gordon, eds., *The Economics of New Goods* (Chicago: University of Chicago Press, 1997).

[28] Parente, S. L., "Technology Adoption, Learning-by-Doing, and Economic Growth," *Journal of Economic Theory* 63 (1994), 346–369.

[29] Redding, S., "The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D," *Economic Journal* 106 (1996), 458–470.

[30] Rhee, Y. W., B. Ross-Larson, and G. Pursell, *Korea's Competitive Edge* (Baltimore: The Johns Hopkins University Press, 1984).

[31] Rogers, E. M., *Diffusion of Innovations,* 4th Edition (New York: Free Press, 1995).

[32] Rosenberg, N., *Inside the Black Box: Technology and Economics* (Cambridge: Cambridge University Press, 1982).

[33] Schultz, T. W., "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature* 13 (1975), 827–846.

[34] Solow, R. M., "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics* 39 (1957), 312–320.

[35] Stokey, N. L., "Learning by Doing and the Introduction of New Goods," *Journal of Political Economy* 96 (1988), 701–17.

[36] Stokey, N. L. and R. E. Lucas, Jr. with E. C. Prescott, *Recursive Methods in Economic Dynamics* (Cambridge: Harvard University Press, 1989).

[37] Stoneman, P. and N. J. Ireland, "The Role of Supply Factors in the Diffusion of New Process Technology," *Economic Journal* 93 (1983), 66–78.

[38] Welch, F., "Education in Production," *Journal of Political Economy* 78 (1970), 35–59.

[39] Young, A., "Invention and Bounded Learning by Doing," *Journal of Political Economy* 101 (1993), 443–472.

[40] Zeckhauser, R., "Optimality in a World of Progress and Learning," *Review of Economic Studies* 35 (1968), 363–365.