Rosenberg's "Learning by Using" and Technology Diffusion*

Toshihiko Mukoyama†
Department of Economics
Concordia University
and
CIREQ

August 2004

Abstract

This paper formulates Rosenberg's (1982) "learning by using" as a stochastic process. The producer of machines learns from the experience of users. Due to this learning, the quality of machines improves over time. It turns out that the process of this improvement approximately takes an exponential form. This improvement process, combined with the growth of demand due to the improvement, can produce an S-shape diffusion curve of machines. Strong demand and advancement of communication technology increase the diffusion speed. The distributional property of the stochastic process and the implications for inequality across machine users are also explored.

Keywords: Learning by Using, Technology Diffusion

JEL Classifications: O31, O33

---

*Part of this research was circulated under the title "A Theory of Technology Diffusion". I would like to thank Jeremy Greenwood for his guidance and Shakeeb Khan for his help. I would also like to thank two anonymous referees, Mark Bils, Per Krusell, Ayşegül Şahin, and Roxanne Stanoprud for helpful comments. All remaining errors are mine.

†Address: Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, QC H3G 1M8, Canada. Tel: +1-514-848-2424 ext.3927. Fax: +1-514-848-4536. E-mail: mukoyama@alcor.concordia.ca

1 Introduction

This paper explores a stochastic model of learning. Since the innovative work by Arrow (1962), the literature on economic growth has been emphasizing the importance of learning activities in the growth process. In the literature, the main focus has been on a specific form of learning – "learning by doing". Typically, learning by doing is formulated as the productivity gain due to the producer's past production experience.

In this paper, we consider a different form of learning. In our model, the producer learns from users. We are particularly interested in the case of capital goods (machine) producers. For example, newly developed machines may have many "bugs". 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. The importance of this type of learning has been recognized by many economic historians. In particular, Rosenberg (1982) termed this type of learning activity as "learning by using" and 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)

However, this type of learning has not been formulated in theoretical literature. This paper aims to fill this gap.

Von Hippel (1988) argues that users play an important role in improving technology. Through many case studies, such as scientific instruments, semiconductor processes, and PC board assembly, he found that users often make crucial contributions to the improvement of technology. The case of electron microscopes (p.17) provides a clear-cut example of the process emphasized by Rosenberg in the above quote. Von Hippel describes the process of an important improvement in the design of electron microscopes (self-cleaning of microscope aperture). The self-cleaning aperture was first invented by a microscope user at Harvard University. He presented a paper at the Electron Microscope Society of America, and an improved aperture was commercialized by a company that sells electron microscopic supplies, who learned about the method from this user.

MacLeod (1992) also emphasizes this type of interaction between users and capital-goods suppliers. She studies the 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)". The empirical study by Bahk and Gort (1993) suggests the importance of this channel in modern manufacturing. Using the plant-level data during 1973–1986, they show that the industry-wide increases in the stock of knowledge affect output only if the increases are related to embodied technical change in capital stock.

The concept of learning by using bears some similarity to the "second-order learning" by Adler and Clark (1991). They distinguished between first-order learning (learning by direct workers, based on repetition and incremental development of expertise) and second-order learning (learning due to explicit managerial or engineering action to change the technology, the equipment, the process, or the human capital). In a case study of an electronic equipment company, they found that second-order learning is important and is largely induced by the production experience. They also found that the capital-intensive production area derives as much learning as the labor/materials-intensive production area, which suggests the importance of learning in capital. Their result is consistent with the plant-level evidence provided by Bahk and Gort (1993), which shows that a new plant experiences a productivity increase from capital stock for five or six years after birth.

In economic development literature, the effect called "learning by exporting" is often pointed out as an important source of productivity gain. When goods are exported from developing countries to more developed countries, the purchasing agents often suggest ways to improve the quality of the goods or the efficiency of the manufacturing process. For example, it is often argued that learning by exporting has played an important role in Korea's economic development [Rhee, Ross-Larson, and Pursell (1984, Chapter 4)]. Learning by exporting can be interpreted as a particular type of learning by using.

As Rosenberg emphasizes, learning by using is even more important in modern high-technology industries. For example, many computer software companies distribute "beta-versions" of new softwares before their formal release, to facilitate this type of learning.

The work by Jovanovic and Nyarko (1995) is similar to ours in spirit. They construct a stochastic model to provide a microfoundation of Arrow's learning by doing. We attempt to do the same for Rosenberg's learning by using. Jovanovic and Nyarko consider the case where the result of learning is embodied in the skills of the workers, while we consider the case where it is embodied in the capital goods.

We show that the diffusion curve of a new technology can exhibit an S-shape under reasonable assumptions. Our theory contributes to the recent theoretical literature which tries to explain S-shaped diffusion curves from a solid microfoundation.[1]

As Jovanovic and Nyarko emphasize, one of the benefits of building an explicit stochastic model is that it enables us to examine distributional consequences. We also explore the distributional implications of our model.

In what follows, we will employ a statistical model to formulate the idea of learning by using. The next section presents the model and characterizes the law of motion for the machine quality. We also compare our learning process to the existing learning functions in the literature. It is shown that the basic model can be extended to the case where task assignment is correlated across users. In Section 3, we analyze the learning process further by numerical simulations. We show that under reasonable assumptions the diffusion curves

[1] See, for example, Jovanovic and Lach (1989), Chari and Hopenhayn (1991), and Götz (1999).

exhibit an S-shape. In Section 4, we characterize the stochastic properties of the learning process. In Section 5, we study the implications of inequality across machine users. Section 6 concludes.

2 Model

2.1 Basic Model

Consider a producer of machines selling to many users. Each machine can be used for K tasks. The operation of each task may be subject to "bugs". For each task i and period t ($t = 0, 1, 2, ...$), define a variable $m_{i,t}$ by

$$m_{i,t} \equiv \begin{cases} 1 & \text{if task } i \text{ has a bug at period } t, \\ 0 & \text{if it does not.} \end{cases}$$

There are Z_t users who operate machines in each period t. Each user employs his machine to carry out one of these tasks, chosen at random. Here, this random assignment of tasks is assumed to be independent across time and across users. At time t, the operation of $M_t \equiv \sum_{i=1}^{K} m_{i,t}$ tasks are subject to "bugs". Thus, a user is hit by an error (finds an error) with probability M_t/K. If a user finds an error, it is reported to the producer[2] in the end of the period, and it is fixed immediately. (From that time, that particular task becomes error-free.) Since the assignment of tasks is random, two or more users may be assigned to the same task and find the same error at the same time. If this duplication occurs, only one error is fixed (since the error is found in one task). When X_t non-duplicated errors are found at time t, the number of tasks that are subject to bugs drops to $M_{t+1} = M_t - X_t$ at time $t + 1$.

Denote $\theta_t \equiv M_t/K$ and $z_t \equiv Z_t/K$. The machine is "better" when θ_t is small. z_t measures the amount of machine use in period t. It is expected that, on average, θ_{t+1} will become smaller when z_t is larger. The next proposition shows that it is in fact the case.

[2] For example, we can consider an arrangement that the producer pays some fee for each reported error.

Proposition 1 *Given θ_t, the expected ratio of remaining errors in a machine at period $t+1$, $E[\theta_{t+1}|\theta_t]$, follows*

$$E[\theta_{t+1}|\theta_t] = \theta_t \left(1 - \frac{1}{K}\right)^{z_t K}. \tag{1}$$

Proof: For each tasks that contains an error, the probability that the error is not found (no one is assigned to that task) is

$$\Pr\{m_{i,t+1} = 1 | m_{i,t} = 1\} = \left(1 - \frac{1}{K}\right)^{Z_t}.$$

Since there are M_t tasks that are subject to bugs, the expected number of tasks that contains error after Z_t users used the machine is:

$$E[M_{t+1}|M_t] = M_t \left(1 - \frac{1}{K}\right)^{Z_t}.$$

Dividing both sides by K (and using $Z_t = z_t K$) yields (1). □

The behavior of θ_t becomes deterministic when K is large.

Proposition 2 *When we take a limit $K \to \infty$ keeping $z_t = Z_t/K$ constant, the ratio of remaining errors in a machine at period $t+1$, θ_{t+1}, converges to*

$$\theta_{t+1} = \theta_t e^{-z_t}. \tag{2}$$

Proof: Note that

$$\lim_{K\to\infty} \left(1 - \frac{1}{K}\right)^{z_t K} = e^{-z_t}.$$

M_{t+1} can be regarded as a sum of M_t i.i.d. random variables which takes 1 with probability $(1 - 1/K)^{z_t K}$ and 0 with probability $1 - (1 - 1/K)^{z_t K}$. Therefore, from the Law of Large Numbers, θ_{t+1} converges to $\theta_t e^{-z_t}$ in probability. □

The equation (2) shows that machines are improved at a rate which is an exponential function of the amount of users.

The "learning function" (2) is quite intuitive. The marginal rate of learning at $z_t = 0$ is

$$\left.\frac{\partial \theta_{t+1}}{\partial z_t}\right|_{z_t=0} = -\theta_t.$$

This corresponds to the fact that the probability of finding an error for the first user is θ_t. As z_t increases, the rate of improvement exhibits "decreasing returns": $\partial \theta_{t+1}/\partial z_t$ becomes smaller in absolute value. This reflects the duplication effect in the error-finding process: as z_t increases, it becomes more likely that a user finds an error that has already been found by another user in the same period.

The functional form of (2) ensures that z people using the machines for two periods leads to the same degree of improvement as if $2z$ people were using the machines for one period. This transpires since $(\theta_t \cdot e^{-z}) \cdot e^{-z} = \theta_t \cdot e^{-2z}$. This is because the duplication does not depend on a particular time frame (since the task assignment is i.i.d.).

2.2 Comparison to Other Learning Functions

Our learning process is comparable to the learning functions used in the literature. In the literature, many learning functions are formulated in terms of productivity. In our context, it is natural to assume that productivity is a decreasing function of θ_t, since the machines with lower θ_t are better. (Consider, for example, the case where all the learning takes place in one firm or one plant.) Specifically, assume that productivity Q_t can be represented as $Q_t \equiv 1 - \theta_t$. Then, (2) can be rewritten as

$$Q_{t+1} - Q_t = (1 - e^{-z_t}) \cdot (1 - Q_t).$$

When z is exogenous and constant over time, $(1 - e^{-z_t})$ can be regarded as a (positive) constant parameter. Then, this formulation becomes equivalent to the learning function of Parente (1994, 2000).[3] Our model provides a microfoundation for this learning function, although the context is slightly different from Parente's.

[3] Jovanovic and Nyarko (1995) state that this functional form has been popularly used in economics and psychology.

Another popular approach in the literature is to formulate productivity as a function of the cumulative production experience. Assuming that z_t can be a proxy for production experience in each period, our formulation results in

$$Q_{t+1} = 1 - \theta_0 e^{-\Lambda_t}, \tag{3}$$

where $\Lambda_t \equiv \sum_{s=0}^{t} z_s$ is the cumulative experience (by users). Arrow's formulation is

$$Q_{t+1} = \alpha_1 \Lambda_t^{\alpha_2}, \quad \alpha_1, \alpha_2 > 0,$$

where the producer's cumulative experience is Λ_t.[4] Jovanovic and Nyarko's learning process produces the average productivity $E[Q_{t+1}]$ that follows

$$E[Q_{t+1}] = 1 - \frac{1}{\varpi_1 + \varpi_2 \Lambda_t}, \quad \varpi_1, \varpi_2 > 0.$$

Our learning function (3) suggests an alternative functional form in the case of "learning by using". A notable common property between our learning function and Jovanovic and Nyarko's is that learning is bounded. Young (1991, 1993) utilizes an exponential learning function similar to (3). Young argues that assuming bounded learning is reasonable considering the evidence on plateauing and the stagnation in premodern history.

2.3 Extending to Correlated Task Assignment

In the basic model in Section 2.1, we assumed that the tasks are assigned independently across users. In reality, however, it is typically the case that some tasks are preferred over others by many users. In this situation, the probability of finding an error is not uniform across tasks. Here, we extend the basic setup to analyze the learning process when user's actions are correlated. We consider two different extensions.

[4] Arrow used cumulative investment as Λ_t here.

2.3.1 Limited Task Assignment Model

Assume that, in the beginning of each period, a subset of the total tasks is randomly selected. Call these tasks "preferred tasks".[5] Each user is assigned a task randomly from these preferred tasks. Suppose that there are k preferred tasks. Clearly, $1 \leq k \leq K$ has to be satisfied. When $k = 1$, there is a perfect correlation among users' activities: all the users are assigned the same task. The case of $k = K$ boils down to our basic model.

Let $p \equiv k/K$. For a particular task which has a bug, the probability that it is a preferred task is p. When it is a preferred task, the probability that the error is not found is $(1-1/k)^{Z_t}$. With probability $(1-p)$, nobody is assigned to the task and the error is not found. Therefore, the probability that an error is not found from a particular task at period t is

$$\Pr\{m_{i,t+1} = 1 | m_{i,t} = 1\} = p\left(1 - \frac{1}{k}\right)^{Z_t} + (1-p).$$

Since there are M_t such tasks, the expected number of tasks in the next period is

$$E[M_{t+1}|M_t] = M_t\left[p\left(1 - \frac{1}{k}\right)^{Z_t} + (1-p)\right]. \tag{4}$$

Therefore, the counterpart of (1) is

$$E[\theta_{t+1}|\theta_t] = \theta_t\left[p\left(1 - \frac{1}{k}\right)^{\frac{z_t k}{p}} + (1-p)\right]. \tag{5}$$

Note that when $k = 1$ (perfect correlation), (4) becomes

$$E[M_{t+1}|M_t] = M_t - \frac{M_t}{K}.$$

In this case, everyone is assigned the same task. One error is found with the probability M_t/K.

[5] This implies that the set of preferred tasks changes over time. If the preferred tasks are the same over time, the analysis is much simpler. This extension has the same effect as reducing K, which implies a smaller z_t when Z_t is the same. This will make the error-finding more efficient within the preferred tasks since user-task ratio is higher. The error-finding rate from the entire tasks is lower, since some tasks are never tried.

From (5), the counterpart of (2) can be calculated as (taking limit $K \to \infty$ with keeping z_t and p constant)

$$\theta_{t+1} = \theta_t \left[pe^{-\frac{z_t}{p}} + (1-p) \right]. \tag{6}$$

In Appendix A, it is shown that the right-hand-side of (2) is smaller than the right-hand-side of (6), that is,

$$e^{-z_t} \leq pe^{-\frac{z_t}{p}} + (1-p). \tag{7}$$

Moreover, the difference between them is decreasing in p, which implies that the higher correlation in task assignment makes the learning slower. The intuition is simple: higher correlation in task assignment makes the duplication effect more severe.

2.3.2 Sequential Task Assignment Model

Alternatively, consider a setting where tasks are assigned sequentially to the users within a period. The first user is assigned to a task randomly from the whole set of tasks. With probability $(1-q)$, the second user is assigned to the same task as the first user. With probability q, the second user is assigned to a task randomly (from the whole set of tasks), in the same way as the first user. In general, the ith user ($i \geq 2$) is assigned to the same task as the $(i-1)$th user with probability $(1-q)$, and he is assigned to his task randomly with probability q. Here, $q \in [0,1]$ measures the degree of correlation. When q is small, the correlation is large.

Appendix B shows that the expected number of errors in the next period is

$$E[M_{t+1}|M_t] = M_t \left(1 - \frac{q}{K}\right)^{Z_t - 1} \left(1 - \frac{1}{K}\right). \tag{8}$$

When $q = 0$ (perfect correlation), (8) becomes

$$E[M_{t+1}|M_t] = M_t - \frac{M_t}{K}.$$

When $q = 1$, the model is equivalent to the basic model. The counterpart of (2) can be calculated as

$$\theta_{t+1} = \theta_t e^{-z_t q}. \tag{9}$$

Clearly, $e^{-z_t} \leq e^{-z_t q}$ and the efficiency of learning is higher when q is larger (correlation is lower). The intuition is the same as the previous section.

3 Deterministic Properties of θ_t

In this section (and parts of the following two sections), we make some parametric assumptions and numerically analyze the properties of the basic model. In this section, we focus on the learning process (2). This exercise will uncover the behavior of θ_t when K is large.

3.1 Constant z_t

First, consider the case where the number of users in each period (z_t) is constant at z. Then, (2) becomes

$$\theta_{t+1} = \theta_t \cdot e^{-z} = \theta_0 \cdot e^{-z(t+1)}$$

for any t. Figure 1 plots the dynamics of θ_t for $z = 0.2$ and $z = 0.4$, starting from $\theta_0 = 1$.

[Figure 1 Here]

3.2 Time-Varying z_t and Diffusion Curves

It is more reasonable to assume that z_t changes as time passes. When θ_t is large, not many users would want to use the machine. As θ_t becomes smaller, more and more users will start using the machine. In fact, this is the diffusion process that Rosenberg (1982) emphasizes.[6] He writes:

> The diffusion process is usually dependent upon a stream of improvements in the performance characteristics of an invention, its progressive modification and

[6] Olmstead (1975) argues that the rapid adoption of leapers and mowers in the mid-1850s (20 years after their invention) is largely due to the improvement in their design, which increased machine longevity, versatility, and productivity, and reduced the risks and uncertainty of breakdowns.

adaptation to suit the needs or specialized requirements of various submarkets, and upon the availability and introduction of other complementary inputs that make an original invention more useful. (p.21)

Here, we assume that

$$z_t = (1 - \kappa\theta_t)z_0, \tag{10}$$

where κ and z_0 are given constant. z_0 is the maximum value that z_t can attain, and can be interpreted as the potential strength of demand. In Appendix C, it is shown that (10) can be derived from a model of heterogenous users. In Figure 2, we plot the process of θ_t assuming $\theta_0 = 1$, $\kappa = 0.95$, and $z_0 = 0.4$, 0.6, and 0.8. Larger z_0 implies more feedback, and naturally θ_t declines faster when z_0 is larger.

[Figure 2 Here]

[Figure 3 Here]

Figure 3 shows a time path of engine maintenance expense, taken from Rosenberg (1982). Assuming that the maintenance expense is proportional to θ_t, this time path fits well with the time path of θ_t in Figure 2. The decline of maintenance cost is slow initially (in fact, it even increases[7]), accelerates for a while, and then flattens out.

[Figure 4 Here]

In Figure 4, we draw the curves representing z_t/z_0. These curves represent the amount of users (relative to z_0), and can be interpreted as diffusion curves of the machine. They exhibit an S-shape. This S-shape pattern arises in the diffusion of many products and technologies.[8] Figure 5 exhibits the empirical diffusion curves for various products.

[Figure 5 Here]

[7] Rosenberg (1982) explains: "The rise of maintenance costs during the first year of introduction reflects the impact of early design problems that were not anticipated prior to the rigors of actual on-line operations (p.131)". This story is consistent with our model.

[8] For recent surveys on technology diffusion, see Karshenas and Stoneman (1995), Geroski (2000), and Hall and Khan (2002).

Why are the diffusion curves S-shaped in Figure 4? The change in z_t can be calculated as

$$\begin{aligned} z_{t+1} - z_t &= (1 - \kappa\theta_{t+1})z_0 - (1 - \kappa\theta_t)z_0 \\ &= \left(1 - \kappa\theta_t e^{-z_t}\right) z_0 - (1 - \kappa\theta_t)z_0 \\ &= \left(1 - \kappa\theta_t e^{-(1-\kappa\theta_t)z_0}\right) z_0 - (1 - \kappa\theta_t)z_0 \\ &= \left[1 - e^{-(1-\kappa\theta_t)z_0}\right] \kappa z_0 \theta_t. \end{aligned}$$

The first part, $[1 - e^{-(1-\kappa\theta_t)z_0}]$, is decreasing in θ_t. This is because the number of users at time t, $(1 - \kappa\theta_t)z_0$, is small when θ_t is large. Therefore the amount of improvement is small when θ_t is small. The second part, $\kappa z_0 \theta_t$, is increasing in θ_t. As θ_t declines, this part becomes smaller. Intuitively, finding the last few bugs is difficult, even if there are large number of users. In the examples of Figure 4, the effect of first part dominates when θ_t is close to 1, and the speed of diffusion is slow initially. The diffusion speed increases as θ_t declines, but when θ_t becomes sufficiently small, the second effect starts to dominate, and the diffusion slows down. More formally, define the slope of the diffusion curve (multiplied by z_0) at t as

$$F(\theta_t) \equiv \left[1 - e^{-(1-\kappa\theta_t)z_0}\right] \kappa z_0 \theta_t.$$

and the change in the slope at t as

$$G(\theta_t) \equiv F(\theta_{t+1}) - F(\theta_t) = F\left(\theta_t e^{-(1-\kappa\theta_t)z_0}\right) - F(\theta_t).$$

Since $d\theta_t/dt < 0$ and $\lim_{t\to\infty} \theta_t = 0$, a set of sufficient conditions for the S-shaped diffusion curve is that $G(\theta_t)$ is positive for large values of θ_t and negative for small values of θ_t.

In Figure 4, different z_0 create different diffusion curves. Initially, the interaction between z_t and θ_t magnifies the difference in diffusion speed among the cases with different z_0. The diffusion curves "converge" as they reach the ceiling.

[Figure 6 Here]

Larger z_0 can be interpreted as the strength of demand. Griliches (1957) shows that the diffusion speed of hybrid corn is faster in the areas where average corn acres per farm is larger. Figure 6 shows the diffusion curves from Griliches (1957).

[Table 1 Here]

Larger z_0 can also be interpreted as larger feedback from users. Therefore, z_0 can be increased by the advancement of information and communication technology. Table 1 shows how many years it took for major inventions to diffuse to 25% of the population. Clearly, this duration is becoming shorter in recent years. In the context of our model, this phenomenon can be explained by the recent information technology revolution – more efficient communication and information transmission increased z_0.[9]

4 Stochastic Properties of θ_t

Jovanovic and Nyarko (1995) argued that one advantage of explicitly analyzing the stochastic process of learning is that we can obtain insight on more than only the first moment of the learning process. There are two stochastic (distributional) aspects in our model. First, due to the stochastic nature of the learning process, the actual learning speed may differ even if we start from the same initial conditions. Therefore, the realized values of θ_t may differ across different capital goods, even when all the economic environments are the same. Second, due to the randomness in task assignment, users may be assigned to tasks with or without bugs. If the production performances of users differ by whether or not a bug exists in the assigned task, there are distributional consequences across users. In this section, we analyze the stochastic properties of θ_t, and in the next section we explore the distributional properties across users.

When K becomes large, the Central Limit Theorem ensures that (given θ_t) the behavior of $\sqrt{K}(\theta_{t+1} - E[\theta_{t+1}|\theta_t])$ can be approximated by a normal distribution.

Proposition 3 *Let*

$$\mu_t \equiv E[\theta_{t+1}|\theta_t] = \theta_t \left(1 - \frac{1}{K}\right)^{z_t K}$$

[9] Cooley and Yorukoglu (2002) also argue that the arrival of an "information age" accelerates technology diffusion. In their model, the diffusion of new goods become faster due to increased efficiency in the production of "information input" (knowledge and information contained in products).

and

$$\sigma_t \equiv \sqrt{\left(1-\frac{1}{K}\right)^{z_tK}\left[1-\left(1-\frac{1}{K}\right)^{z_tK}\right]}.$$

Then, as $K \to \infty$, $\sqrt{K}(\theta_{t+1}-\mu_t)$ converges in distribution to a random variable which follows a normal distribution $\mathcal{N}(0, \theta_t\sigma_t^2)$.

Proof: Denote $m_{i,t+1}$, $i = 1, ..., M_t$, as a random variable which attains 0 if a bug is found at time t and 1 if not. All $m_{i,t+1}$ (conditional on information at time t) are i.i.d. with mean

$$\psi_t \equiv \left(1-\frac{1}{K}\right)^{z_tK}$$

and variance

$$\sigma_t^2 = \left(1-\frac{1}{K}\right)^{z_tK}\left[1-\left(1-\frac{1}{K}\right)^{z_tK}\right].$$

Clearly, $M_{t+1} = \sum_{i=1}^{M_t} m_{i,t+1}$. Hence,

$$M_{t+1} = M_t\psi_t = M_t\left(1-\frac{1}{K}\right)^{z_tK}.$$

Now, consider the behavior of $M_{t+1} = \sum_{i=1}^{M_t} m_{i,t+1}$ as $M_t \to \infty$. Since $m_{i,t+1}$ are i.i.d., we can apply the Lindberg-Levy Central Limit Theorem[10] to conclude that

$$\sqrt{M_t}\left(\frac{1}{M_t}\sum_{i=1}^{M_t} m_{i,t+1} - \psi_t\right) = \sqrt{M_t}\left(\frac{M_{t+1}}{M_t} - \psi_t\right) \xrightarrow{d} \mathcal{N}(0, \sigma_t^2).$$

Since

$$\sqrt{M_t}\left(\frac{M_{t+1}}{M_t} - \psi_t\right) = \sqrt{\frac{K}{\theta_t}}(\theta_{t+1} - \mu_t),$$

the proposition follows. □

It is difficult to analytically characterize the behavior of θ_t when K is small. In the following, we attempt to characterize the distributional properties when K is small by Monte-Carlo simulation.

[10]This particular case of binomial distribution is also called De Moivre-Laplace Theorem [see, e.g., Billingsley (1995, p.358)].

4.1 Constant z_t

First, we simulate the case with constant z_t. We assume that $K = 50$ and $Z_t = 10$. Therefore, this case corresponds to the case with $z_t = 0.2$ in the previous section. In Figure 7, the solid line represents the average value of θ_t. (This approximately corresponds to the solid line in Figure 1.) The dashed lines represent the top 10% and bottom 10% of the distribution.[11] Figure 8 shows the evolution of the distribution of θ_t as histograms (size of a bin = 0.02). It can be seen that the dispersion of θ_t increases initially, and then declines as θ_t starts to approach zero.

[**Figure 7 Here**]

[**Figure 8 Here**]

4.2 Time-Varying z_t

Next, we examine the case where z_t depends on θ_t. We assume that Z_t follows the following equation

$$Z_t = (1 - \kappa\theta_t)Z_0, \tag{11}$$

therefore $z_t = (1 - \kappa\theta_t)z_0$. We assume that $\theta_0 = 1$, $\kappa = 0.95$, $K = 50$, and $Z_0 = 30$. In terms of equation (10), this corresponds to the case where $z_0 = Z_0/K = 0.6$. Since Z_t has to be an integer, we used the largest integer that does not exceed the right-hand-side value of (11).[12] Figure 9 shows the mean, top 10%, and bottom 10% of the distribution. Figure 10 shows the evolution of the distribution as histograms (size of a bin = 0.02). It exhibits a similar pattern to the previous example — the dispersion increases initially, and then declines as θ_t reaches close to zero.

[**Figure 9 Here**]

[**Figure 10 Here**]

[11] We simulated the process 10,000 times.

[12] Therefore, the average rate of decline in θ_t is somewhat smaller than the simulation with $z_0 = 0.6$ in the last section.

5 Distributions Across Users

5.1 Distributions Across One-Task Users

In this section, we consider the cross-sectional distribution across users, each of whom performs one task. A user is randomly assigned one task in each period. Assume that user i's production $y_{i,t}$ is

$$y_{i,t} \equiv \begin{cases} 1 & \text{if the assigned task doesn't have a bug,} \\ 0 & \text{if the assigned task has a bug.} \end{cases} \tag{12}$$

At time t, a user is assigned a task that contains a bug with probability θ_t, which implies that

$$y_{i,t} = \begin{cases} 1 & \text{with probability } (1-\theta_t), \\ 0 & \text{with probability } \theta_t. \end{cases}$$

Therefore, when Z_t is large enough, the cross-sectional variance of $y_{i,t}$ across users becomes $\theta_t(1-\theta_t)$. When $\theta_0 > 0.5$, the dispersion increases initially, and then decreases. This pattern is similar to the dynamics of inequality in the model of Jovanovic and Nyarko (1995). In particular, if $\theta_0 > 0.5$ and θ_0 is not too large, then the period of the increasing inequality is relatively short. This is consistent with the plant-level evidence in Bahk and Gort (1993, Table 4), who observed that the adjusted R^2 of the production function estimation first falls and then rises as the plants age. If $\theta_0 < 0.5$, the heterogeneity falls monotonically. This pattern is observed by Griliches and Regev (1995) [cited by Jovanovic and Nyarko (1995)].[13] In Griliches and Regev's samples, efficiencies are less heterogenous among older firms.

5.2 Distributions Across Multiple-Task Firms

Now, suppose that the economy consists of firms who hire many users (workers). There are N firms in the economy, each of which consists of L users. Therefore $Z_t = LN$ for all t. Firm

[13] If there is heterogeneity in productivity across users and the diffusion is endogenous (as in the model of Appendix C), inequality may not decline as θ_t becomes small, since new users with different productivity are added as θ_t falls. However, if θ_t represents a skill requirement for operating a machine and the users are heterogenous in terms of skills (but productivity is the same as long as a user can operate the machine), inequality tends to decline as θ_t approaches zero. See Mukoyama (2004) for such a model.

j assigns a task to each user randomly from the set $\mathcal{S}_j \subseteq \mathcal{K}$, where $\mathcal{K}$ is the set of the entire tasks. Assume that $\mathcal{S}_j$ is time-invariant, the same size across firms, and $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$ for $i \neq j$.

Let the number of elements in $\mathcal{S}_j$ be S and the number of bugs in $\mathcal{S}_j$ at time t be $\chi_{j,t}$. Also define $\phi_{j,t} \equiv \chi_{j,t}/S$ and $l \equiv L/S$. Then, the stochastic behavior of $\phi_{j,t}$ can be analyzed in the same manner as the behavior of θ_t in Section 4. In particular, as $S \to \infty$ (keeping l constant), $\sqrt{S}(\phi_{j,t+1} - \varphi_{j,t})$ converges to a normal distribution with mean 0 and variance $\phi_{j,t}\omega_t^2$, where

$$\varphi_{j,t} \equiv \phi_{j,t} \left(1 - \frac{1}{S}\right)^{lS}$$

and

$$\omega_t \equiv \sqrt{\left(1 - \frac{1}{S}\right)^{lS} \left[1 - \left(1 - \frac{1}{S}\right)^{lS}\right]}.$$

Since all the firms are symmetric, the cross-sectional distribution of $\phi_{j,t}$ behaves in the same way as the distribution of θ_t in Figures 7 and 8.

How is the behavior of $\phi_{j,t}$ translated into the behavior of output in each firm? Below, we consider several scenarios for how the tasks are aggregated to the output of the firm.

First, we consider a production structure called the "series" structure in Jovanovic and Nyarko (1995):

$$Y_{j,t} = \min_{i \in \mathcal{L}_j} \{y_{i,t}\},$$

where $Y_{j,t}$ is the output of firm j at time t, $\mathcal{L}_j$ is the set of users at firm j, and $y_{i,t}$ is defined in (12). This structure expresses the idea that the tasks are complementary inputs in the production process. With this production structure, $Y_{j,t} \in \{0, 1\}$, and the probability that $Y_{j,t} = 1$ is equal to the probability that all $y_{i,t}$ is 1. Therefore,

$$\Pr[Y_{j,t} = 1] = (1 - \phi_{j,t})^L . \tag{13}$$

The polar opposite of the series structure is called the "parallel" structure, which is formulated as

$$Y_{j,t} = \max_{i \in \mathcal{L}_j} \{y_{i,t}\}.$$

In this case,

$$\Pr[Y_{j,t} = 1] = 1 - (\phi_{j,t})^L. \tag{14}$$

Figure 11 plots the relationships (13) and (14) when $L = 5$. This pattern emerges for any $L > 1$. With the "series" production structure, the probability becomes sensitive to the change in $\phi_{i,j}$ when $\phi_{i,j}$ is close to zero. With the "parallel" structure, the change in probability is large when $\phi_{i,j}$ is close to 1. Thus, the parallel structure tends to create an inequality in the early stage of learning, while the series structure gives rise to an inequality in the late stage of learning.

[Figure 11 Here]

As an alternative setting, suppose that the production structure is additive:

$$Y_{j,t} = \sum_{i \in \mathcal{L}_j} y_{i,t}. \tag{15}$$

Since $y_{i,t}$ is i.i.d. (given $\phi_{j,t}$), the Central Limit Theorem ensures that $Y_{j,t}$ tends to follow a normal distribution (given $\phi_{j,t}$) when L is large. If, instead, the production structure is multiplicative:[14]

$$Y_{j,t} = \prod_{i \in \mathcal{L}_j} \exp(y_{i,t}), \tag{16}$$

$Y_{j,t}$ tends to follow a lognormal distribution when L is large,[15] since

$$\log(Y_{j,t}) = \sum_{i \in \mathcal{L}_j} y_{i,t}. \tag{17}$$

In this case, the distribution of output across firms is skewed to the right even when $\phi_{j,t}$ is distributed symmetrically across j.[16]

[Figure 12 Here]

Figure 12 plots the result of Monte-Carlo simulation in the case of additive production structure (15). It plots the distribution of the output $Y_{j,t}$ across firms at $t = 1$, $t = 3$, and

[14] Jovanovic and Nyarko (1995) use this type of production structure. See also Beckmann (1977), Rosen (1982), and Kremer (1993).

[15] If we multiply $y_{i,t}$ together instead of $\exp(y_{i,t})$, this boils down to the "series" production structure above.

[16] It is well known that income distribution and firm size distribution tend to exhibit a right skew.

$t = 10$. The parameter values are: $S = 50$, $L = 10$, $\chi_{j,0} = 50$ for all j, and $N = 10,000$. It can be seen that inequality first rises, and then falls.[17] From (17), Figure 12 can be viewed as the plot for the multiplicative case (16), by reinterpreting that the horizontal axis is for $\log(Y_{j,t})$ instead of $Y_{j,t}$.

6 Conclusion

This paper formulated Rosenberg's (1982) "learning by using" as a stochastic process. The producer of machines learns from the experience of users. Due to this learning effect, the quality of machines improves over time. We showed that the process of this improvement can be approximated by an exponential form. This improvement process, combined with the growth of demand due to improvement, can produce an S-shape diffusion curve of machines. Strong demand and advancement of communication technology increase the diffusion speed.

The distributional property of the stochastic process is also explored. We found that when the initial quality of the machine is low, the dispersion of machine quality tends to increase first, and to decline as the machines diffuse. It is also shown that the cross-sectional distribution of output across users tends to follow a similar pattern. An important future research topic is to empirically examine our predictions utilizing micro-level data.

[17] As can be seen from Figure 12, the distribution tends to become left-skewed as t becomes larger. This counter-factual prediction can be remedied by assuming that a innovation of entirely new machine can occur when the current technology is sufficiently improved. Mukoyama (2004) proposes such a model of innovation.

Appendix

A Inequality (7)

First note that (7) holds with equality when $p = 1$. Therefore, it suffices to show that

$$f(p) \equiv e^{-z_t} - [pe^{-\frac{z_t}{p}} + (1 - p)]$$

is decreasing in p. Since

$$f'(p) = \left(1 + \frac{z_t}{p}\right) e^{-\frac{z_t}{p}} - 1,$$

to show that $f'(p) \leq 0$, it suffices to show that

$$\left(1 + \frac{z_t}{p}\right) e^{-\frac{z_t}{p}} \leq 1.$$

Taking logs on both sides and denoting $x \equiv z_t/p$, this is equivalent to

$$\log(1 + x) - x \leq 0.$$

It is clear that this holds for any $x \geq 0$.

B Derivation of Equation (8)

Denote the number of the errors not found before the ith user by M^i. M^i is a random variable. By definition, $M^1 = M_t$. Let the expected probability that the ith user finds an error be π_i. Clearly, $\pi_1 = M_t/K$. It is also clear that for $i \geq 2$

$$\pi_i = q \cdot \frac{E[M^i]}{K} + (1 - q) \cdot 0,$$

where $E[\cdot]$ is the expectation taken in the beginning of the period.

Let the event $A_i \equiv \{i\text{th user finds an error}\}$ and $A_i^c \equiv \{i\text{th user does not find an error}\}$. The following holds from the law of iterated expectations:

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

Clearly the first term is $-\pi_{i-1}$ (since one error is found if A_{i-1} happens), and the second term is $E[M^{i-1}]$ (since nothing changes from the past user if the error is not found). Then, noting that $\pi_i = qE[M^i]/K$ for $i \geq 2$ and $\pi_1 = M_t/K$,

$$
\begin{aligned}
E[M^i] &= E[M^{i-1}]\left(1 - \frac{q}{K}\right) \\
&= E[M^2]\left(1 - \frac{q}{K}\right)^{i-2} \\
&= M_t\left(1 - \frac{q}{K}\right)^{i-2}\left(1 - \frac{1}{K}\right).
\end{aligned}
$$

Setting $i = Z_t + 1$ (since "after Z_t users" is equivalent to "before the $(Z_t + 1)$th user") yields (8).

C Heterogenous Users Model

Suppose that the production function of a user is

$$y_t = \lambda(1 - \kappa\theta_t),$$

where $\kappa \in (0, 1]$ is a constant. Assume that users are heterogenous in λ. Let the rental price of the capital be γ. Then, a user decides to operate if and only if

$$\lambda(1 - \kappa\theta_t) \geq \gamma,$$

That is,

$$\lambda \geq \frac{\gamma}{1 - \kappa\theta_t}.$$

Assume that λ follows a Pareto distribution with distribution function

$$D(x) = 1 - \frac{b}{x},$$

where b is a parameter and $x \geq b$. Then, the number of users who actively operate is (we assume that $\gamma \geq b$):

$$z_t = 1 - D\left(\frac{\gamma}{1 - \kappa\theta_t}\right) = \frac{b}{\gamma}(1 - \kappa\theta_t).$$

By denoting $z_0 \equiv b/\gamma$, we obtain (10).

References

[1] Adler, P.S., Clark, K.B., 1991. Behind the Learning Curve: A Sketch of the Learning Process, Management Science 37, 267-281.

[2] Arrow, K. J., 1962. The Economic Implications of Learning by Doing, Review of Economic Studies 29, 155-173.

[3] Bahk, B.-H, Gort, M., 1993, Decomposing Learning by Doing in New Plants, Journal of Political Economy 101, 561-583.

[4] Beckmann, M., 1977, Management Production Functions and the Theory of the Firm, Journal of Economic Theory 14, 1–18.

[5] Billingsley, P., 1995, Probablility and Measure, Third Edition (John Wiley & Suns, New York).

[6] Chari, V.V., Hopenhayn, H., 1991. Vintage Human Capital, Growth, and the Diffusion of New Technology, Journal of Political Economy 99, 1142-1165.

[7] Cooley, T.F., Yorukoglu, M. 2002. Some Macroeconomic Implications of an Information Age, mimeo. New York University and the University of Chicago.

[8] Federal Reserve Bank of Dallas, 1996. The Economy at Light Speed, Annual Report.

[9] Geroski, P. A., 2000. Models of Technology Diffusion, Research Policy 29, 603-625.

[10] Götz, G., 1999. Monopolistic Competition and the Diffusion of New Technology, RAND Journal of Economics 30, 679-693.

[11] Griliches, Z., 1957. Hybrid Corn: An Exploration in Economics of Technological Change, Econometrica 25, 501-522.

[12] Griliches, Z., Regev, H., 1995. Firm Productivity in Israeli Industry 1979–1988, Journal of Econometrics 65, 175-203.

[13] Hall, B. H., Khan, B., 2002. Adoption of New Technology, mimeo. University of California at Berkeley.

[14] Jovanovic, B., Lach, S., 1989. Entry, Exit, and Diffusion with Learning by Doing, American Economic Review 79, 690-99.

[15] Jovanovic, B., Nyarko, Y., 1995. A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves, Brookings Papers: Microeconomics, 247-305.

[16] Karshenas, M., Stoneman, P., 1995. Technological Diffusion, In: P. Stoneman (eds.), Handbook of the Economics of Innovation and Technological Change, Blackwell, Oxford: Blackwell.

[17] Kremer, M., 1993. The O-Ring Theory of Economic Development, Quarterly Journal of Economics 108, 551-575.

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

[19] Mukoyama, T., 2004. Diffusion and Innovation of New Technologies under Skill Heterogeneity, mimeo. Concordia University.

[20] Olmstead, A. L., 1975. The Mechanization of Reaping and Mowing in American Agriculture, 1833-1870, Journal of Economic History 35, 327-352.

[21] Parente, S. L., 1994. Technology Adoption, Learning-by-Doing, and Economic Growth, Journal of Economic Theory 63, 346-369.

[22] Parente, S. L., 2000. Learning-by-Using and the Switch to Better Machines, Review of Economic Dynamics 3, 675-703.

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

[24] Rosen, S., 1982. Authority, Control, and the Distribution of Earnings, Bell Journal of Economics 13, 311-323.

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

[26] von Hippel, E., 1988. The Sources of Innovation (Oxford University Press, Oxford).

[27] Young, A., 1991. Learning by Doing and the Dynamic Effects of International Trade, Quarterly Journal of Economics 106, 369-405.

[28] Young, A., 1993. Invention and Bounded Learning by Doing, Journal of Political Economy 101, 443-472.

Tables

	Year Invented	Years to Spread
Electricity	1873	46
Telephone	1876	35
Automobile	1886	55
Airplane	1903	64
Radio	1906	22
Television	1926	26
VCR	1952	34
Microwave Oven	1953	30
PC	1975	16
Cellular Phone	1983	13
Internet	1991	7

Table 1: Spread of products to a quarter of the population
Source: Federal Reserve Bank of Dallas (1996)

Figures

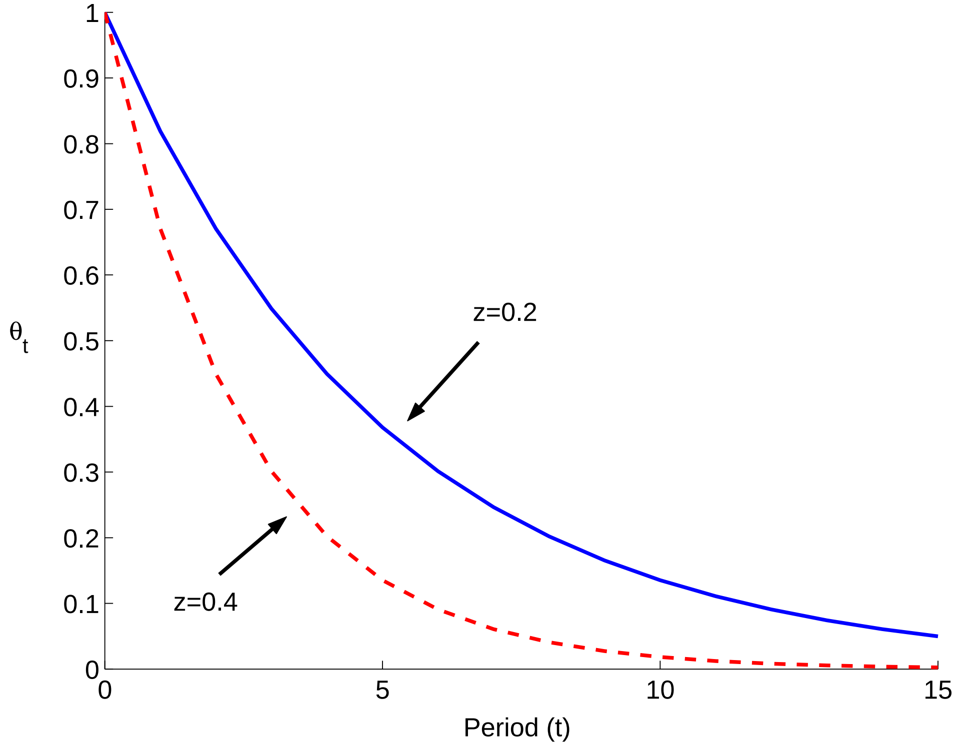

Figure 1: θ_t when $z = 0.2$ and $z = 0.4$.

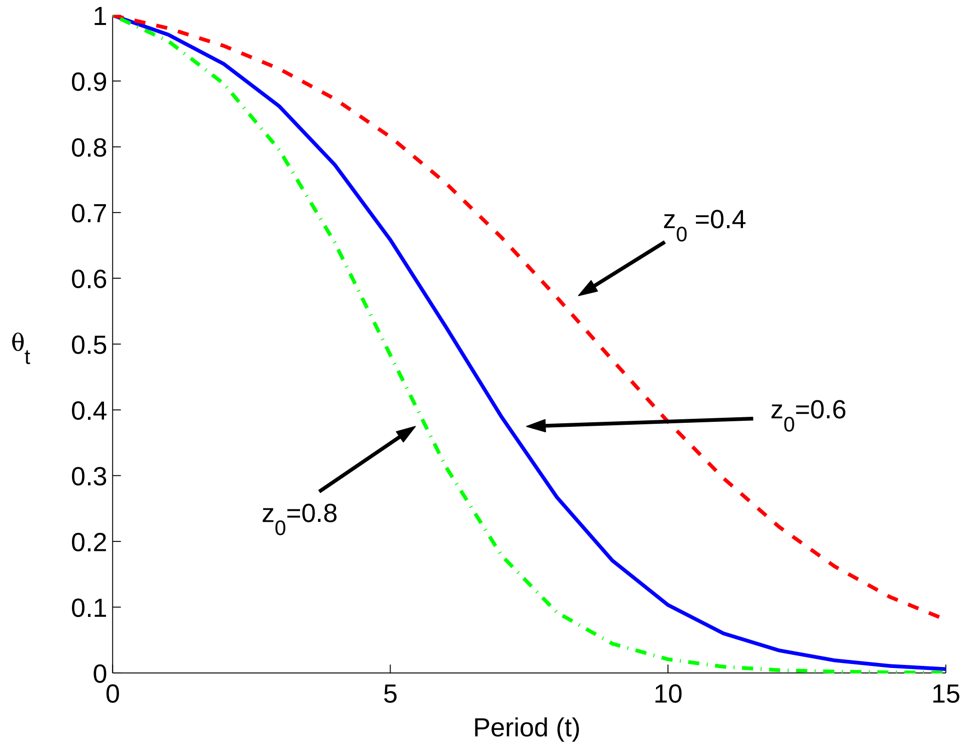

Figure 2: θ_t when $z_0 = 0.4$, 0.6, and 0.8.

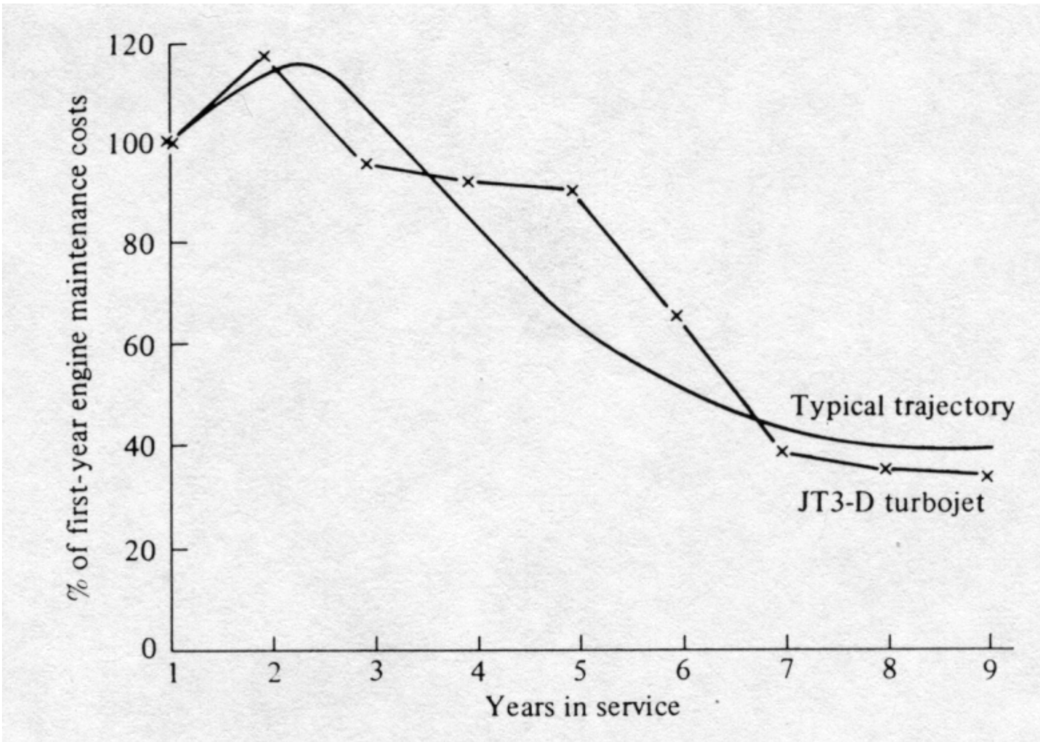

Figure 3: Engine maintenance expense
Source: Rosenberg (1982)

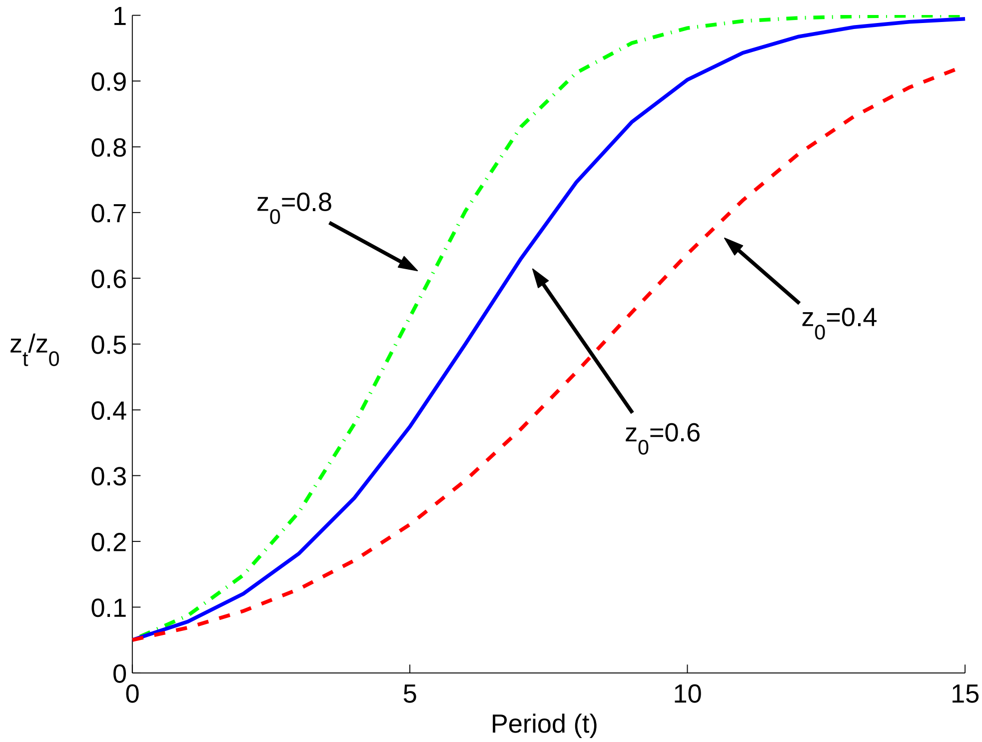

Figure 4: z_t/z_0 when $z_0 = 0.4$, 0.6, and 0.8.

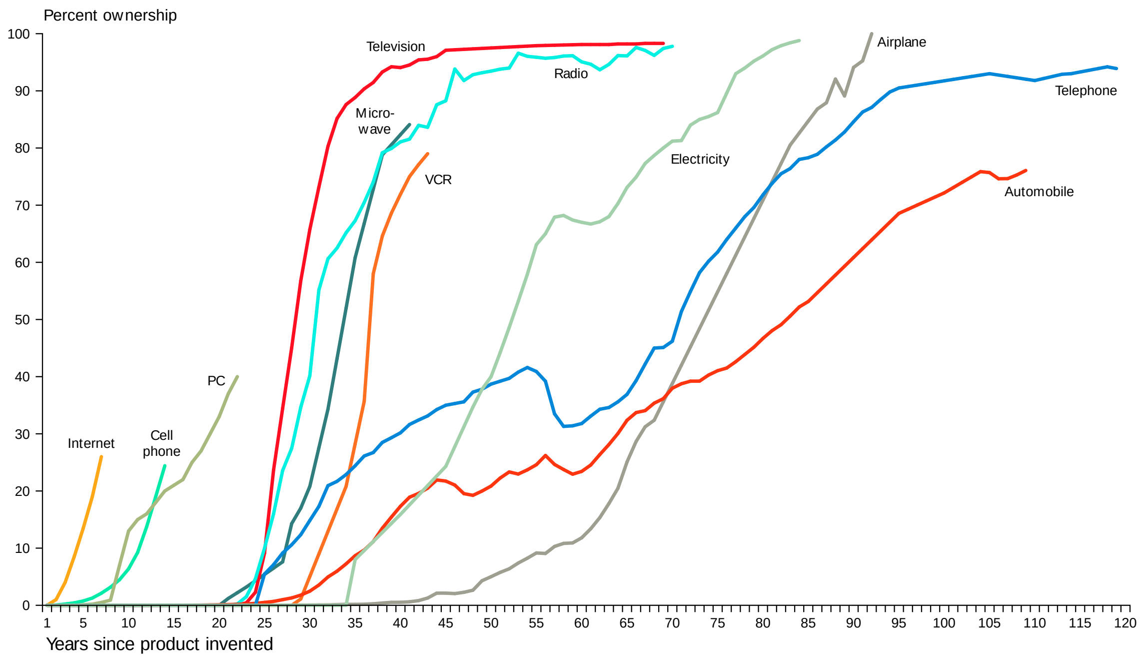

Figure 5: Diffusion curves

Source: Federal Reserve Bank of Dallas (1996)

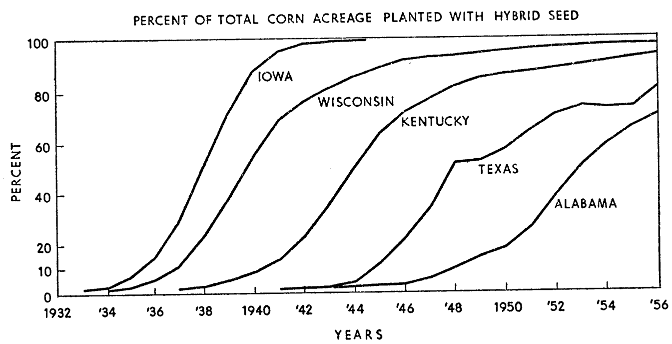

Figure 6: Percentage of total corn acreage planted with hybrid seed
Source: Griliches (1957)

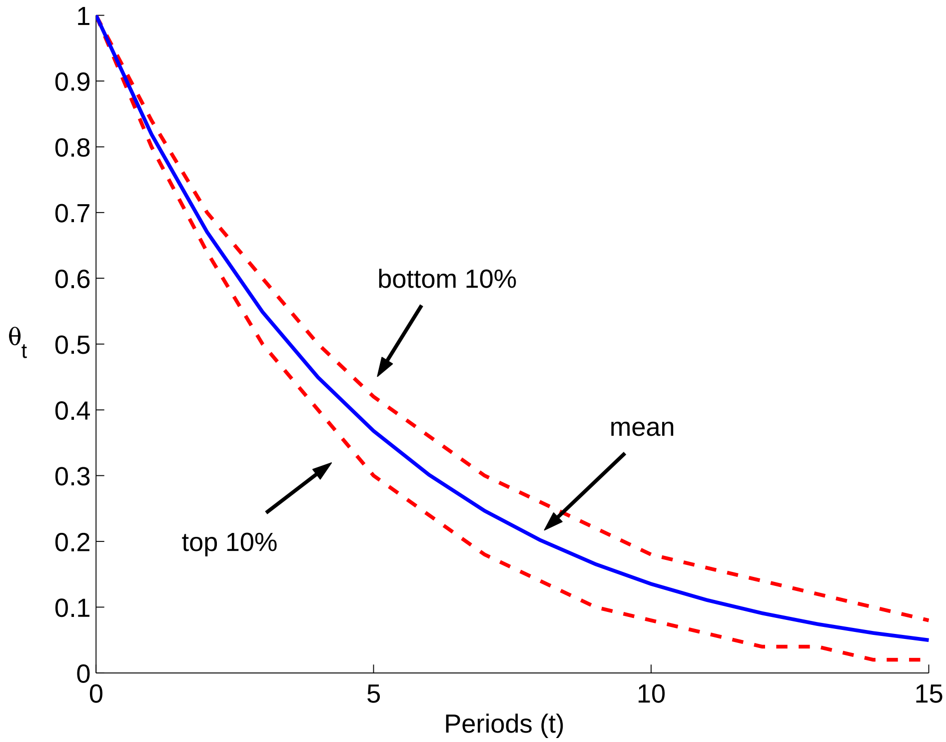

Figure 7: Paths of mean θ_t, top 10%, and bottom 10%

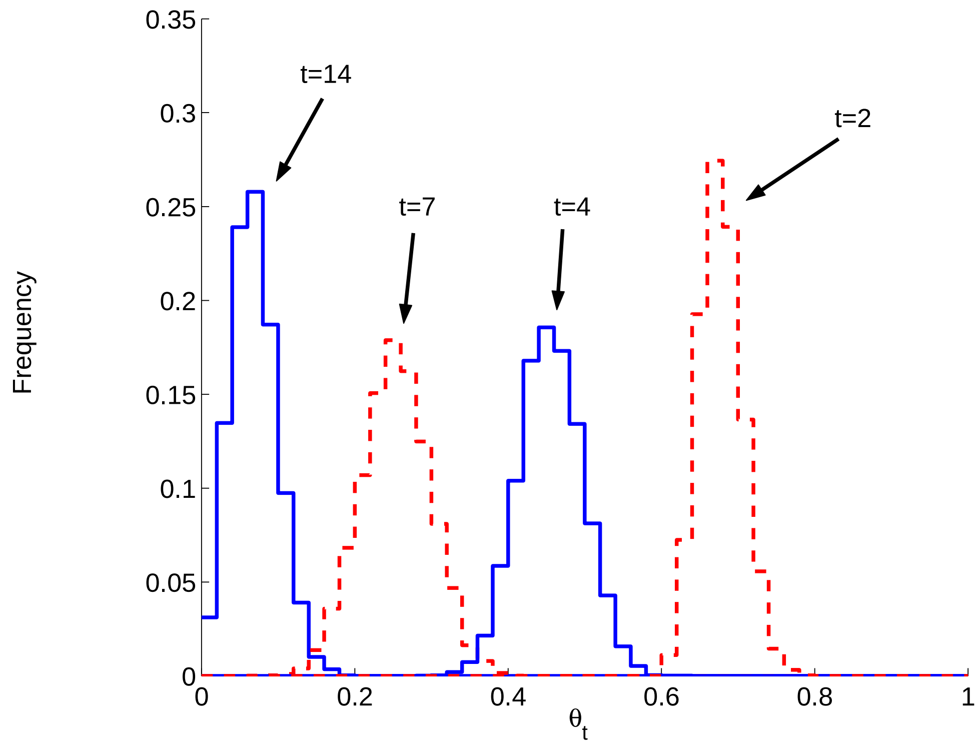

Figure 8: Distributions of θ_t

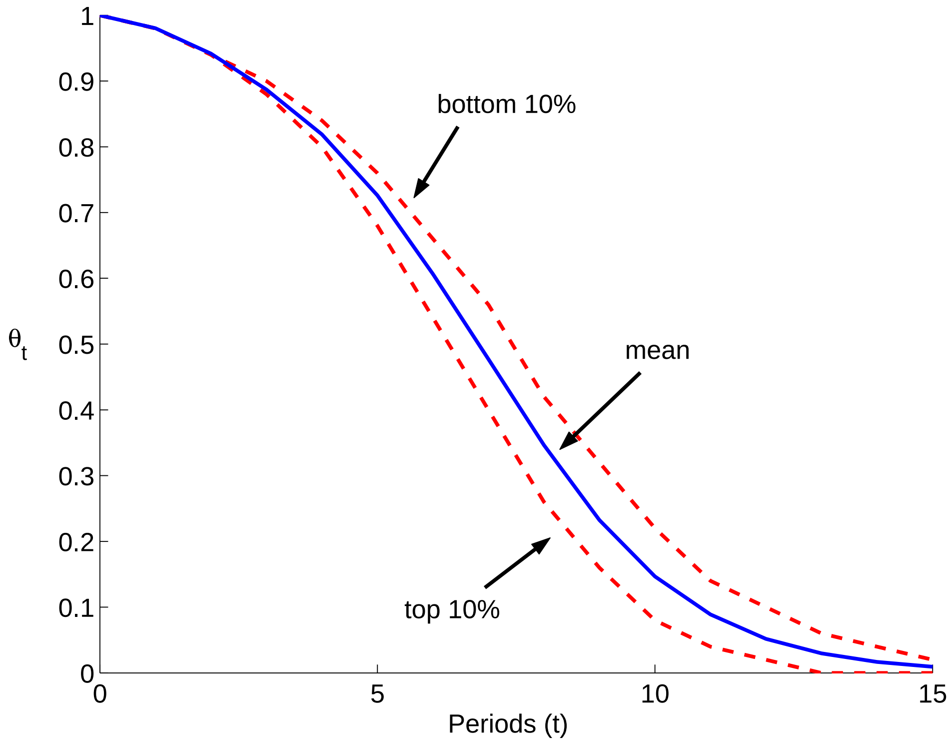

Figure 9: Paths of mean θ_t, top 10%, and bottom 10%

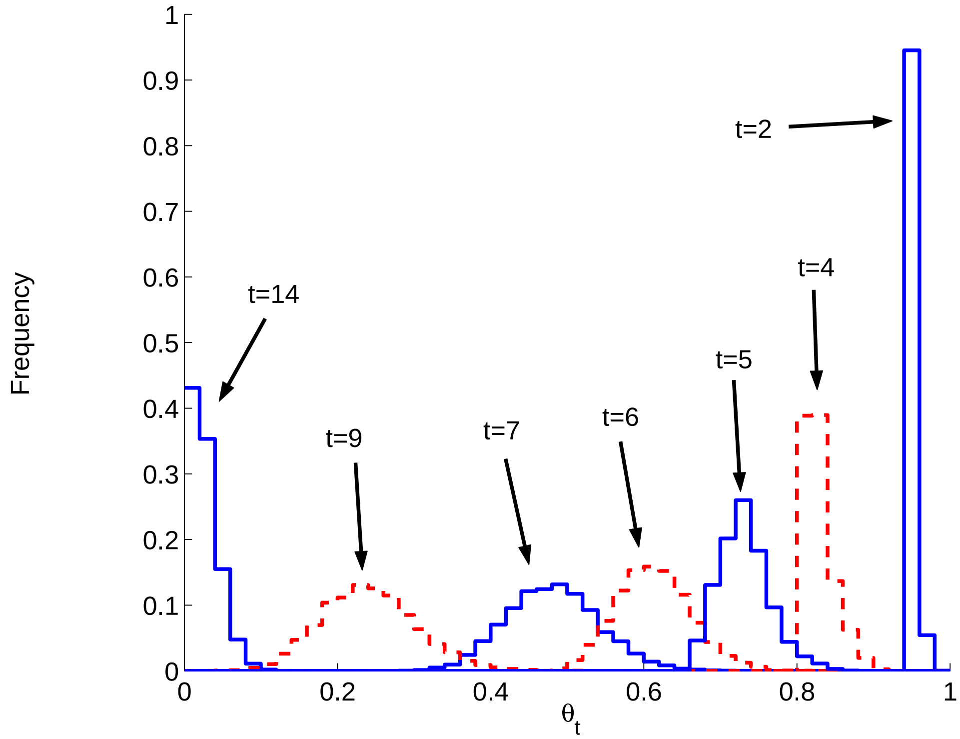

Figure 10: Distributions of θ_t

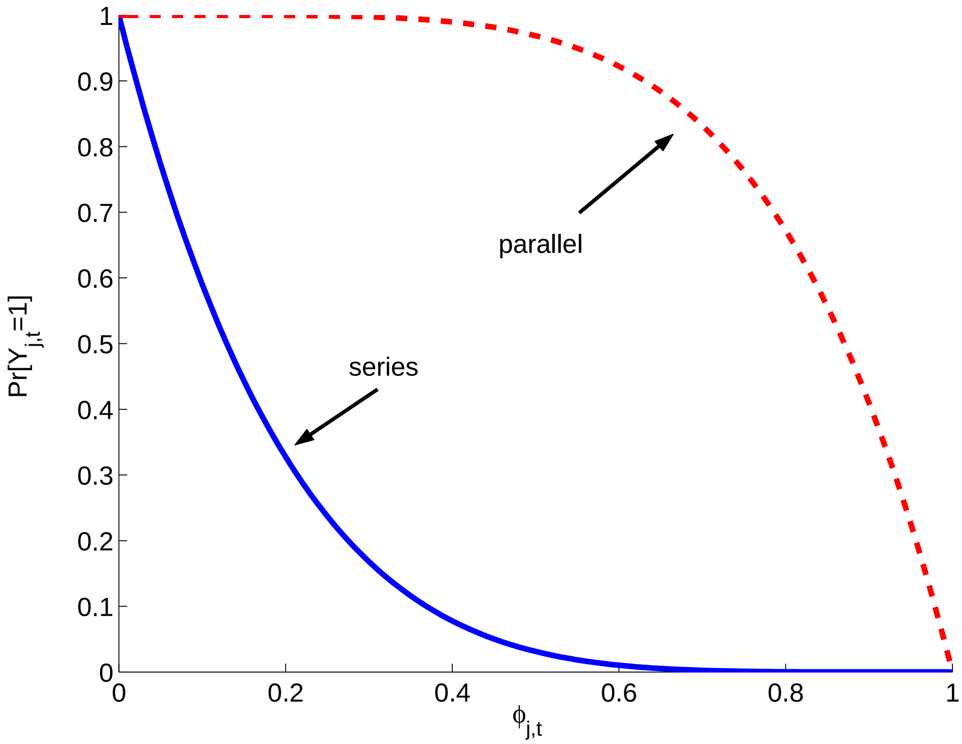

Figure 11: $\Pr[Y_{j,t} = 1]$ given $\phi_{j,t}$

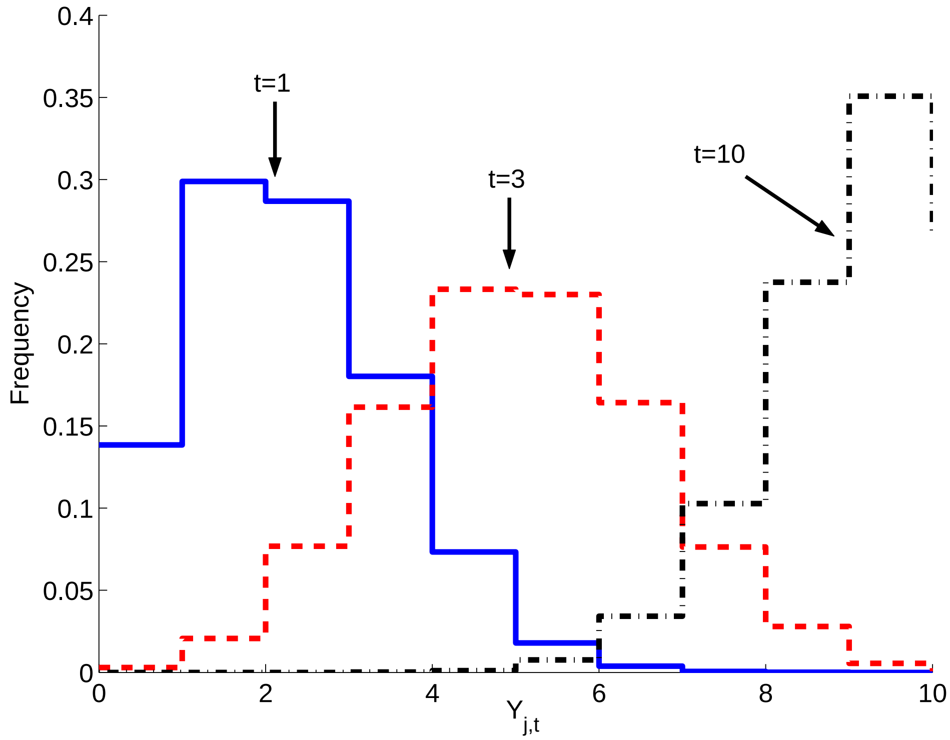

Figure 12: Distributions of $Y_{j,t}$