**BOYAN JOVANOVIC**
*New York University*

**YAW NYARKO**
*New York University*

# A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves

WHERE DOES TECHNOLOGICAL progress come from and what determines its rate of advance? In answering these questions, it is useful to decompose technological progress into the invention of new techniques and products and the improvement of existing ones.

Roughly speaking, the economist sees invention as the result of research and development, and improvement as the result of experience—learning by doing. Because productivity growth on any single process is likely to be bounded, invention is the origin of long-run productivity growth. But the "level" effects of improvement on productivity have, in some activities, been found to be huge—on the order of several *hundreds* of percentage points. Thus understanding how the process of improvement works will help us better account for growth. This paper concerns itself with a simple model of *one* of the forces involved in improvement, namely, the improvement in productive efficiency that occurs as a joint product with output, or learning by doing.

**The Learning Curve**

Formal study of the learning curve began in 1899 when two psychologists, Bryan and Harter, published their celebrated study on adult skill learning. The activity they studied was the sending and receiving of Morse code, and they found that after ten months' experience, the operators were four or five times as productive as in their first month. Since then, economists, industrial engineers, and psychologists have found many other instances of productivity growth associated with the accumulation of experience. The gains vary from a few percentage points in some activities carried out by individuals to tenfold increases in group activities such as the assembly of airframes.[1]

Of course, much of what may look like learning by doing is in fact the result of costly investments. These investments are concentrated at the outset when a new process or product is introduced—training the work force in the initial stages of production, allocating extra engineers at the early stages, getting the newly installed equipment up and running—and they have a lasting effect on productivity. Nevertheless, the sheer volume of examples from the real world, added to the even greater number of experimental studies, leaves a strong impression that learning by doing matters a lot.

*Our Approach*

When a new production method is set up in a plant, and when unit costs decline with experience, where is the increased efficiency "embodied"? Is it the firm's management that makes better operating and purchasing decisions, or the firm's labor force that gets better, or the firm's capital stock that is being debugged? Because new products and processes differ widely in their capital intensity, their complexity, and so on, the answer will depend on the case at hand. Nevertheless, these different sources of productivity improvement all have one thing in common: As experience accumulates, someone—the manager, the worker, the engineer, the head of purchasing—makes better decisions. Thus on an abstract level, productivity growth can be seen as the result

---

1. Bryan and Harter (1899), Middleton (1945), Baloff (1966), and Argotte and Epple (1990).

of a better decision that solves some optimizing problem. This is the approach that we shall take.

In our model, ''skill'' is knowledge. Each period, someone—say, the worker—must make a decision. With each repetition of the activity, the worker grows more informed and the decision gets better and better. Hence the model generates a learning curve—a positive relation between experience and productivity. With its emphasis on optimal learning and optimal decisions, our model differs from the main body of the existing literature on this relationship. In light of this, the reader may well ask the following two questions: Is this a plausible way to model productivity growth on a new activity, and does information theory add anything, or can a simpler, noninformational approach do the job as well?

### Plausibility

Can a formal representation of knowledge plausibly describe skill? The answer depends on the type of skill in question. The Bryan-Harter study suggests a productivity breakdown into three basic skills:

- Language skills—needed to learn the Morse code,
- Perceptual skills—needed to identify patterns of dots and dashes, and
- Motor skills—needed to manipulate the sending key.[2]

Because there is an ideal way to send and receive telegraph signals that allows us to define a *best* decision, the development of each of these three skills was essential for narrowing the gap between actual decisions and the best decision. We assume that actual decisions are optimal in that they reflect information received to date. But optimal decisions do not coincide with the best decision, simply because the worker has not yet learned what that best decision is. Even as a matter of descriptive realism, then, the worker could be viewed as having initial beliefs about the meaning of the code and of the dot-dash patterns, and that he learns their correct meaning in a Bayesian fashion, by updating his initial

---

2. Ackerman (1988) argues, in an article well known to psychologists, that performance on a task is driven initially by ''general intelligence,'' then by ''perceptual speed,'' and eventually by ''psychomotor ability,'' a decomposition similar to the three basic skills outlined here.

beliefs in light of the evidence, and so improves his performance. This seems an appropriate way to model the types of language and perceptual skills described above. The development of motor skills, however, is another matter; although the above argument may apply in the telegraph example, elsewhere we must use an "as if" argument. A laborer gets better with experience partly because working makes him fitter and stronger, not necessarily because he is using logic or processing information. When this is the case, we hope that the growth in productivity will behave *as if* the worker's actions kept improving because of mounting evidence about what kind of action was best.

### What Information Theory Adds

Why use information theory explicitly to get a learning curve? Why not proceed by simply assuming a deterministic learning curve with a functional form that is flexible enough to fit a variety of situations? This approach has proved useful in both economics and psychology,[3] and a popular functional form is $dq(t)/dt = b[c - q(t)]$, $q(0)$ given, where $q(t)$ is a measure of an agent's efficiency or productivity on a technology that he has used for $t$ periods, and where $b$ and $c$ are constants. This simple functional form captures two essential properties of learning: its total *scope*, as measured, say, by the "progress ratio" $c/q(0)$—the maximal, eventual efficiency, divided by the initial efficiency, and its *speed*, as measured by $b$. Why, then, explore a more complicated structure?

We can think of two advantages of our decision-theoretic approach. First, the above equation specifies the first moments only—mean productivity grows as a function of experience. But in fact, learning has something to do with variability of productivity, even among equally experienced decisionmakers. For one reason or another, some agents learn faster than others, and when the learning curve is steep, differences among agents are potentially larger. Workers, plants, and firms will have different productivities because some have learned a production method faster than others. What is the precise relation between the scope of learning and the variance of efficiency in a cohort of decisionmakers who are initially identical? Our model supplies a precise answer

---

3. Parente (1994) and Eyring, Johnson, and Francis (1993) are examples in economics and psychology where this functional form is fruitfully explored.

to this question. It also has something to say about the skewness in the distribution of efficiencies. In other words, the learning curve, the variance of efficiency, and the skewness of efficiency are all related to a few common parameters.

The second advantage of the decision-theoretic approach is that the parameters governing the scope and speed of learning also determine the *switching cost*, expressed in the loss of know-how when a familiar method of production is abandoned in favor of a new, untried one. Because the parameters $b$, $c$, and $q(0)$ are not the outcome of an underlying informational structure, the deterministic equation displayed above does not suggest an appropriate value for the switching cost, which therefore must be specified exogenously, as a free parameter.[4] And yet we know that the amount of human capital lost to switching should depend on how much was accumulated. Again, our model gives a precise answer to this question.

### Learning and the Complexity of a Decision Problem

In fitting the model to actual learning curves, we encountered the problem that productivity growth in a couple of cases seemed too slow compared with how fast the decisionmaker receives new information. A farmer may take several years to learn a new agricultural method, but this is to be expected because he receives information slowly—in effect, the farmer gets just one piece of information each time the crop is grown, and this happens at most two or three times a year. But on other jobs information comes in at a much higher rate—for instance (and this is a case we analyze) a surgeon that does several operations a week. As we shall see, a surgeon's skills in performing an operation keep improving even after he has performed hundreds of them! A simple decision problem with a single unknown parameter and serially uncorrelated signals does not adequately describe the learning on such an activity, because the model would predict that learning should occur too quickly.

How then does one construct a model showing slow productivity growth in an environment where learning is optimal and behavior is rational? In view of the usual statistical findings that autocorrelation in

---

4. Parente (1994) and Klenow (1993) formulate their models in this way. Jovanovic and Nyarko (1994) analyze optimal switching in the model at hand.

the noise reduces the signal-to-noise ratio in data, our first stab at a solution was to allow the signal on the unknown parameter to be auto-correlated. To our surprise, we found that learning the parameter in question slowed, but not the rate of productivity growth because auto-correlation can make certain types of forecasting easier. Thus some other way was needed to slow down productivity growth.

A natural way to obtain slower learning in this class of models is to increase the number of unknown parameters that the decisionmaker must know to make a good decision. Our way of enlarging the number of parameters also increases the number of decisions that must be taken for the activity to be carried out. Although we do not favor this inter-pretation, we could associate each decision with the exercise of a dif-ferent kind of underlying ability, as in our discussion of the three basic skills above, in which case some activities require the exercise of a greater range of abilities. We think of a problem with a larger number of decisions as "more complex." In short, then, productivity growth is slower when the activity is complex. Indeed, we show that marginal returns to information can be *rising* in such an activity and that its learning curve can then have a convex segment.

### Our Results

Our modeling produced three main findings. First, when we compare the estimation results for the twelve different learning processes on which we have data, the following pattern emerges: processes on which learning is important, as measured by cumulative productivity growth, tend to have longer half-lives of learning, tend to be more complex, and tend to give rise to more inequality in cohorts of learners. Similarly, complex processes have bigger cumulative productivity growth, take longer to learn, and produce more inequality.

Second, and surprisingly, if decisionmakers see somewhat indepen-dent signals about the parameter that they seek to learn, the efficiencies among these agents can vary greatly. Plant managers in the same in-dustry might all be trying to solve the same problem, for example. Thus the model can explain some of the productive efficiency differences among firms or plants—efficiency as measured, say, by total factor productivity (TFP) and documented by Baily and Gersbach in this volume.

Third, if an activity is complex enough, the distribution of efficiencies among decisionmakers is predicted by the model to become skewed to the right. This result is important for the ''frontier production function'' literature, which assumes that the effects of mistakes are to skew the distribution of efficiencies to the left. Even though the underlying shocks in the model are normal, technical inefficiency resulting from mistakes can very well have a right skew, as apparently it does in some industries.

### Previous Theoretical Work

Most work on learning by doing has not used information theory explicitly. Arrow analyzed the case in which the ''doing'' by one firm bestows ''learning'' on all the firms in the economy, so that the resulting benefits are purely external. In contrast, Rosen treated the case in which the benefits of learning by doing are completely internalized. Since then tomes have been written on the subject. Information-theoretic models are few, however. Our model posits that experience allows the decisionmaker gradually to learn the parameters governing the production function on his activity. As he learns, his performance improves. Venezia has modeled productivity growth in exactly this way—although he did not estimate a model and he used least-squares learning instead of the Bayesian method used here.[5] Despite differences in detail, Venezia's message is the same as ours—statistical decision theory is useful in analyzing the productivity-experience relation.

### The Theoretical Framework

We describe the theoretical model in three stages. First we set out the general notions; second, we describe the simplest case of learning just one task on which just a single decision needs to be set; finally, we extend the model to an *N*-task activity.

---

5. Arrow (1962), Rosen (1972), and Venezia (1985). See also Wilson (1975) and Zellner (1971, ch. 11).

*General Notions*

Let $Q$ be the output of some activity, and let the inputs of capital, labor, and materials be $K$, $L$, and $M$. Write the production function as

$$(1) \qquad Q = F(q, K, L, M).$$

The activity may be the production of finished output, or the production of some component; equation 1 can apply to the activity of various entities: the firm, the plant, the team, the manager, the worker. The parameter $q$ measures how efficiently the activity is carried out.

How $q$ enters $F(.)$ depends on the nature of the learning. Bahk and Gort recently tried to find out whether learning at the plant level was mainly organizational or embodied in the inputs.[6] In this notation, their attempt amounted to asking how $q$ enters the production function of the plant. For instance, if organizational learning was mainly by management, one might write $Q = qF(K, L, M)$. And if learning was mainly embodied in labor, one might write $Q = F(K, qL, M)$. Similar adjustments would be made if learning led to an increased efficiency (that is, debugging) of machinery, or if experience led to the finding and purchase of a better quality of material inputs.

We do not take a stand on exactly where $q$ enters or in how many places. We model learning in a way that abstracts from the details of the decision problem at hand. All that we require is a ''best'' way of solving the problem that is not exactly known at the outset. We assume that efficiency, $q$, depends on how closely a decision, $z$, matches an ideal level, or target, $y$:

$$(2) \qquad q = G(y - z).$$

Efficiency is largest when $z = y$. We may think of $(y - z)$ as a mistake that reduces $q$ below its maximal level of $G(0)$. There is no direct cost to learning or to adjusting $z$, only lost output. So $z$ is something other than the quantities $K$, $L$, and $M$. Examples of decisions that $z$ might stand for are the speed of the production line in a factory, the organization of the workspace, the types of machines and raw materials purchased, and the type of labor hired and its assignment to the tasks.

Assume that the function $G(.)$ is known to the decisionmaker. The only unknown is $y$. Under these assumptions, for learning to be gradual,

6. Bahk and Gort (1993, especially table 4).

we must assume that $y$ is random and that the decisionmaker must choose $z$ before seeing $y$. If $y$ were a constant, and if the form of the function were relatively simple—for example, quadratic—the decisionmaker could infer $y$ fairly quickly, and after that, all learning and growth in efficiency would end. So, instead of learning the exact value of $y$, the decisionmaker would be learning about the distribution of $y$. In our model we assume that the distribution of $y$ is known, except for its mean, which we denote by $\theta$.

The decisionmaker learns something about $\theta$ each time that he repeats the activity. As information accumulates, his decisions improve. This learning process produces an upward-sloping learning curve. The curve will always be upward-sloping regardless of how many decisions need to be set, simply because more information is better than less.

Why cannot the agent simply be "told" what $\theta$ is, so that he can calculate the optimal $z$ and solve the problem without actually having to gain experience in production? Indeed, some features of technologies can be explained in manuals and do not have to be learned through trial and error. Other features of the optimal decision are specific to the situation at hand, to the nature of the factors of production, the raw materials, the workspace, the specifications of the output, and so on. These, the decisionmaker will have to infer for himself.

### A One-Task Activity

Suppose an activity requires that only one task be carried out and that just one decision $z$ be taken.[7] Suppose that

$$(3) \qquad q = A[1 - (y - z)^2].$$

The maximal level of $q$, attainable under ideal conditions, is $A$. Each production run $i$ leads to a new value of $y$:

$$(4) \qquad y_i = \theta + w_i,$$

where $i = 0, 1, 2, \ldots$, and where the $w_i$ are independent, normally distributed random variables with mean zero and variance $\sigma_w^2$.

Noise, $w$, is essential if learning and productivity growth are to be gradual. The disturbances $w_i$ represent transitory factors that affect the

---

7. Wilson (1975) and Venezia (1985) discuss learning of such a production function, and Holt and others (1960) were among the first to use a quadratic profit function.

nature of the optimal decision. They are transitory in that they are serially uncorrelated. For example, $\theta$ may represent the optimal way, on average, to perform a surgical operation (indeed, this is the first empirical learning curve that we analyze), but the exact nature of what is done depends on some characteristics of the patient, and these characteristics vary from patient to patient. Or $\theta$ may represent the optimal speed of the production line on an average day. But from day to day, the exact specifications of the output may be slightly different, various members of the work force may be out sick, the raw material inputs may differ in a peculiar way from their average quality, some essential piece of equipment may break down, and so on. This all suggests that transitory effects on the optimal $z$ are present. How transitory they are depends, formally, on the degree of independence of $w$ over production runs. Later, we show that, quite surprisingly, the degree of autocorrelation of the $w$s makes almost no difference to the model. Until then, we shall maintain the assumption that the $w$s are serially independent.

For production run $i$, the decisionmaker must choose his decision $z_i$ *before* seeing $y_i$. Let $\tau$ denote the cumulative number of production runs, $i = 0, 1, 2, \ldots, \tau - 1$. Let $E_\tau(.)$ denote the decisionmaker's expectation of some variable conditional on information that he has seen during the first $\tau$ production runs. Assuming that he is risk-neutral, the decisionmaker will seek to maximize $E_\tau(q_\tau)$ when choosing his decision $z_\tau$ for the $\tau$th production run. The optimal decision is

$$(5) \qquad\qquad z_\tau = E_\tau(y_\tau) = E_\tau(\theta).$$

The second equality in equation 5 follows because $E_\tau(w_\tau) = 0$. This decision is dynamically optimal as well, because the amount of information that the decisionmaker gets does not depend on the value of $z$ that he chooses. Equations 3, 4, and 5 imply the following reduced form:

$$(6) \qquad\qquad q_\tau = A\{1 - [\theta - E_\tau(\theta) + w_\tau]^2 \}.$$

Now let $x_\tau = E_\tau[\theta - E_\tau(\theta)]^2$ denote the posterior variance over $\theta$, given information from the first $\tau$ production runs. If the prior distribution over $\theta$ is normal with variance $\sigma_\theta^2$, then applying Bayes' rule,

$$(7) \qquad\qquad x_\tau = \frac{\sigma_w^2 \sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2 \tau} \ .$$

Then the expected efficiency on production run $\tau$ is

$$(8) \qquad E_\tau\,(q_\tau)\ =\ A(1\ -\ x_\tau\ -\ \sigma_w^2).$$

This is the ''learning curve'' when the activity in question has just one task.

### An N-*Task Activity*

In several of our empirical examples, we encounter rather slow productivity growth. This slow growth cannot be explained by introducing autocorrelation in $w$, but it can be obtained by increasing the number of tasks.

The activity of an organization usually requires that many decisions be made, usually by many individuals. This is as true on an assembly line in an automobile plant, as it is in a firm with a multilayer management structure. Even the activity of an individual worker can be decomposed into more basic ''tasks,'' and, indeed, industrial engineers often divide an activity into stages and time their duration.[8]

Suppose that to complete an activity, $N$ tasks must be performed. The activity might involve a single worker or many. For instance, $N$ may represent the number of stages in an assembly line, with each worker specializing in a single task. In a sense, $N$ measures how complex the activity is.

How should the efficiency of an activity depend on its component tasks? In particular, how essential should each task be to the activity as a whole? In reliability analysis, an activity is said to have a ''series'' structure if it works only when each of its components works. The production function for such an activity is of the Leontief type: $Q = \min(X_1, \ldots, X_N)$, and in it, the $X$s are complements. The polar opposite of this is a ''parallel'' activity that works if any one of its components works. The production function for this activity is $Q = \max(X_1, \ldots, X_N)$, and in it, the $X$s are substitutes. The multiplicative (and hence Cobb-Douglas) functional form we assume below is closer to the Leontief function: each task is essential, but it has more substitution possibilities.[9]

Assume that on the $i$th production run, task $j$ $(j = 1, 2, \ldots, N)$

---

8. Maynard (1971, sec. 3.2) discusses ''time and motion'' studies.
9. Barlow and Proschan (1975) describe reliability analysis.

involves, as before, a decision $z_{ij}$ and a target $y_{ij}$, so that the expression for $q$ on the $i$th production run, is:

$$(9) \qquad q_i = A \prod_{j=1}^{N} [1 - (y_{i,j} - z_{i,j})^2].$$

This is a model in which mistakes are compounded. This type of formulation has been used to study multilayered management organizations and the assignment of heterogeneous workers to organizations.[10] Aside from being easy to analyze, the multiplicative form produces an easy-to-interpret learning curve, and it also generates a right skewness in the distribution of efficiencies, as we explain in the next section. As in the one-task case, assume that

$$(10) \qquad y_{i,j} = \theta_j + w_{i,j},$$

where $j = 1, 2, \ldots, N$.

Although the activity may involve many people, for now we continue to assume that there is just one decisionmaker who sets all the $z$s. The number of tasks, $N$, is exogenous. Of course, the decisionmaker may have chosen $N$ as a solution to a larger problem not described here, but in this case, the following analysis still holds, conditional on the value of $N$ chosen at some prior stage. Assume that for each value of $j$, the agent has a prior distribution over $\theta_j$ that is normal, with variance $\sigma_\theta^2$, and that the $\theta_j$ are mutually independent in the decisionmaker's prior distribution. Assume also that each $w_{i,j}$ is normally distributed, with mean zero and variance $\sigma_w^2$, and that it is independent of the other $w$s.

Together, equations 9 and 10 imply that all tasks are equally important and that the scope and speed of learning are both equal among tasks.[11] This special case involves simple formulas. Now consider the decisionmaker's problem at the start of production run $\tau$. The problem is to choose the decision vector $z_\tau$ ($z_{\tau,1}$, $z_{\tau,2}$, $\ldots$, $z_{\tau,N}$). Assume that the feasible values of $z_\tau$ lie in the region in which the expected value of the contribution of each task is positive. In that case, $E_\tau(q_\tau)$ is strictly concave in the vector of decisions $z$, and the optimal decision is:

10. Beckmann (1977), Rosen (1982), and Kremer (1993).

11. The assumption that the $w_{i,j}$s have the same distribution for each $j$ (so that $y_{i,j}$ is equally informative for all $j$) is an approximation. Industrial engineers have noted that, during the course of an activity, some tasks are done more often than others and are therefore learned faster; see Maynard (1971, pp. 7–104).

(11)                              $z_{\tau, j} = E_\tau (y_{\tau, j}) = E_\tau (\theta_j),$

where $j = 1, 2, \ldots, N$.

During each production run $i$, the decisionmaker observes the vector $(y_{i, 1}, \ldots, y_{i, N})$. When the finished automobile rolls off the line, each of its components—including steering, brakes, tires, suspension, and engine—is checked separately. There is, in other words, no ''signal confusion,'' and the quality of each operation, each task, can be ascertained ex-post.

Because the $\theta_j$ are independent in the decisionmaker's prior beliefs, and because the $w_{ij}$ are independent over $j$, learning is independent over coordinates $j$, in the sense that if $k \neq j$, observing $y_{i, k}$ is of no use for learning $\theta_j$. Having seen the vector $y_{i, 1}, \ldots, y_{i, N}$ for $i = 0, 1, \ldots, \tau - 1$, the decisionmaker's posterior variance over each $\theta_j$ will be $x_\tau$, where $x_\tau$ is given in equation 7. Expected efficiency on the $\tau$th production run will then be

(12)                              $E_\tau (q_\tau) = A(1 - x_\tau - \sigma_w^2)^N.$

This reduced-form relation is the basic outcome of the analysis, and we refer to it as the learning curve. Sometimes we transform it by dividing it by its maximal value, $A(1 - \sigma_w^2)^N$, thus obtaining a curve that is bounded above by unity. Also, we later express this curve as a function of time rather than cumulative production runs.

## Some Properties of the Learning Curve

Before fitting the model to different empirical curves, we discuss a few of the general properties of the model.

### Concavity, Convexity, and ''Plateauing'' of the Learning Curve

In this model, we can easily have the phenomenon of a convex learning curve for a part of its range. That is, the model can generate increasing marginal returns to information. Bryan and Harter found a plateau in the learning curve for receiving telegraphic messages: productivity grew until the sixteenth week, held steady for several weeks and then resumed its growth in the twenty-fourth week, after which sustained growth was observed (productivity doubled during the sub-

sequent sixteen weeks). Such a phenomenon turns out to be consistent with this model.

The following result implies that the larger $N$ is, the more likely a portion of the learning curve will be convex early on:

PROPOSITION 1. *Assume that* $1 - x_\tau - \sigma_w^2 \geq 0$. *Then,* $d^2E_\tau (q_\tau)/d\tau^2 \geq 0$ *if and only if* $x_\tau \geq 2(1 - \sigma_w^2)/(1 + N)$.

Note that when $N = 1$, it is impossible to meet the condition of the proposition without having a nonpositive expected productivity of each task: $1 - x_\tau - \sigma_w^2 \leq 0$. Hence the learning curve must be concave when $N = 1$. Because $x_\tau$ is monotonically decreasing in $\tau$, with $x_0 = \sigma_w^2$, we have

COROLLARY. *The learning curve* $E_\tau(q_\tau)$ *is S-shaped (that is, it has an initial convex portion followed by a concave portion) if* $\sigma_\theta^2 \geq 2(1 - \sigma_\theta^2)/(1 + N)$. *Otherwise it is strictly concave.*

The convex portion occurs initially or not at all. Given the multiplicative functional form of the production function and the symmetry with which the tasks enter the production function, a plateau-like phenomenon (in which learning slows down for a while and then picks up again) is impossible. If both givens are dropped, then getting a plateau is easy. For instance, suppose that the activity still involves $N$ tasks but that the first task affects $q$ additively as follows:

$$(13) \qquad q = A_1 [1 - (y_1 - z_1)^2] + A_2 \prod_{j=2}^{N} [1 - (y_j - z_j)^2]$$

so that if the beliefs and signals are identically and independently distributed as in the case of equation 12,

$$(14) \qquad E_\tau(q_\tau) = A_1 (1 - x_\tau - \sigma_w^2) + A_2 (1 - x_\tau - \sigma_w^2)^{N-1}$$
$$= q^{simple} + q^{complex}$$

The first term on the right-hand side, suggestively defined as $q^{simple}$, is concave in $\tau$. By the corollary to proposition 1, the second term on the right-hand side, defined as $q^{complex}$ because it potentially has many component tasks, is, for $N$ sufficiently large, S-shaped. As figure 1 shows,

if $A_2$ is sufficiently larger than $A_1$, the sum of a simple ($N = 1$) and unimportant (small $A_1$) task and a complex (large $N$), important (large $A_2$) task produces a plateau-like effect on the learning curve.

Figure 1 portrays the function in equation 14 for the case where the prior beliefs over $\theta_1, \ldots, \theta_N$ are identical and mutually independent, with variance $\sigma_\theta^2 = 0.6$; where the $w_1, \ldots, w_N$ are identically and independently distributed with $\sigma_w^2 = 0.3$; and where $N = 50$, $A_1 = 1$, and $A_2 = 10^8$. The large $A_2$ is offset by the large $N$, and as experience $\tau$ accumulates, the expected value of $q$ tends to about 18. As figure 1 clearly shows, a plateau emerges roughly between the first and third trial.

We need to use equation 14 when fitting the model to data on surgical procedures. For all other cases of learning that we consider, the simpler form in equation 12 is adequate.

### Variance in the Distribution of Efficiencies

How much heterogeneity can be explained as an outcome of variations in the rate of learning? In a group of decisionmakers, innately all the same, all with the same level of experience in some activity, some will be expected to have seen more favorable signals—perhaps through luck in trial and error. When the *w*s are, at least to some extent, independent over agents, such chance variations in the realizations of the signals occur, and they, of course, have a greater impact on heterogeneity when the activity being learned offers greater scope for productivity gains. This is why the shape of the learning curve implies something about higher order moments in the distribution of efficiencies. We *shall* now make this precise.

Suppose a group of decisionmakers begins a new activity at the same time. They are all solving the same problem over and over, but they do not share information. What is the variance and skewness of efficiencies among the members of the group, and how do variance and skewness depend on the level of experience? This variance is bigger for complex processes, and it is a nonmonotonic function of experience, first rising, and then falling as experience accumulates.

Suppose that each decisionmaker has the same prior beliefs about $\theta$ and sees a signal in each period as given in equation 4. Suppose, however, that the signals that agents see are mutually independent. This

**Figure 1. A Plateau Occurring When $E(q) = q^{simple} + q^{complex}$**

$E_\tau(q_\tau)$



$\tau$ (Number of experiments)
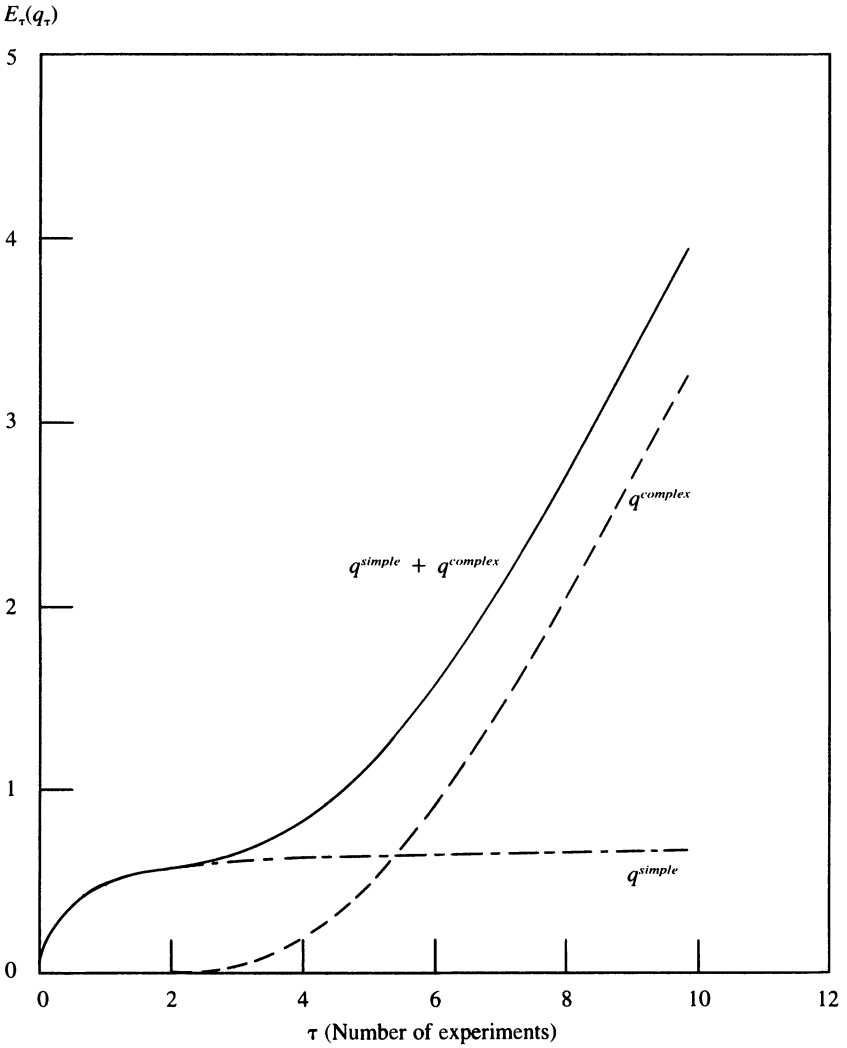
Source: Authors' calculations.

independence of the signals induces differences in $q_\tau$s over agents, and it does so for two reasons. First, because the signal $y$ affects $q_\tau$ directly, even if all agents chose the same $z$, their $(y - z)$s and their efficiencies will be different. And second, as agents see different signals, their beliefs about $\theta$ will temporarily diverge, and they will optimally choose different actions $z$.

We describe the variance of $q_\tau$ by its relative variability, as captured by its squared coefficient of variation, which, of course, equals the variance of $q_\tau$ divided by the square of its mean. We now derive the coefficient of variation for $q$ in the $N$-task case. (The squared coefficient of variation of a random variable is its variance divided by the square of its mean, which also is roughly equal to the variance of the logarithm of the variable in question). Note that in this case, $q$ is a product of $N$ independent random variables. For two independent random variables $\xi_1$ and $\xi_2$ with squared coefficients of variation $v_1$ and $v_2$ respectively, simple algebra shows that the random variable $\xi_1\xi_2$ has a squared coefficient of variation $v_1 + v_2 + v_1v_2$. Induction then produces the result that the squared coefficient of variation of $q$ in the $N$-task case, call it $v(N)$, is related to the squared coefficient of variation of a single task, call it $v$, as follows:

$$(15) \qquad v(N) = (1 + v)^{N-1} - 1.$$

This equation expresses an important feature of complexity:

PROPOSITION 2. *Relative inequality, as measured by v(N), grows exponentially in the level of complexity, N.*

*Any* amount of relative inequality can therefore be obtained from this model, by assuming a production process of sufficient complexity. Of course, the resulting learning curves should also match observed curves—by proposition 1 and its corollary, a large value of $N$ would produce an initially convex portion of the learning curve, but that is seldom seen in practice. So, how much inequality can we realistically explain? To answer this question we need to express $v(N)$, and therefore $v$, in terms of the model's parameters. The first step in this process is to express the variance of efficiency conditional on $\tau$. Because $w$ is normally distributed, $w^2$ is a Chi-squared variate, and its variance is $2\sigma_w^4$. The variable $y - z = w + (\theta - z)$ is also normally distributed

in the population. When $\tau = 0$, everyone chooses the same value of $z$ because everyone has the same priors about $\theta$. When $\tau = \infty$, everyone again chooses the same $z$, because everyone has learned what $\theta$ is. Hence, when $\tau$ is either zero or infinity, the variance of $q$ (conditional on $\theta$) in the group is $2A^2\sigma_w^4$. For intermediate values of $\tau$, the variance is larger, because heterogeneous beliefs over what $\theta$ is cause $z$ to differ among members of the group. The variance of $q_\tau$ for arbitrary $\tau$ is:

$$\text{Var}(q_\tau) = 2A^2\sigma_w^4 \left[ 1 + \frac{\tau\sigma_\theta^2}{(\sigma_w^2 + \tau\sigma_\theta^2)^2} \right].$$

As a function of $\tau$, this expression starts out at $2A^2\sigma_w^4$ when $\tau = 0$, then it rises to a unique maximum, and after that it returns to its original value. It is maximized when $\tau = \sigma_w^2/\sigma_\theta^2$. At this point, $\text{Var}(q_\tau) = 2A^2\sigma_w^4[1 + \sigma_\theta^2/4\sigma_w^2]$.

As $\tau$ rises, the average efficiency in the cohort rises from $A(1 - \sigma_\theta^2 - \sigma_w^2)$ to $A(1 - \sigma_w^2)$. At $\tau = \sigma_w^2/\sigma_\theta^2$, average efficiency is $A(1 - 2\sigma_w^2)$. Hence we obtain the following expressions for $v$ at various levels of experience:

$$(16) \quad v = \begin{cases} \dfrac{2\sigma_w^4}{(1 - \sigma_\theta^2 - \sigma_w^2)^2} & \text{when } \tau = 0, \\[3ex] \dfrac{2\sigma_w^4(1 + \sigma_\theta^2/4\sigma_w^2)}{(1 - 2\sigma_w^2)^2} & \text{when } \tau = \sigma_w^2/\sigma_\theta, \\[3ex] \dfrac{2\sigma_w^4}{(1 - \sigma_w^2)^2} & \text{when } \tau = \infty. \end{cases}$$

Using this result, we can predict the behavior of inequality in the general case by substituting for $v$ into equation 15.

In short then, among a group of decisionmakers who are identical to begin with, inequality should first rise and then decline with experience, and the rise and the decline (as well as the average extent of inequality) should be more pronounced in complex activities.

Our estimates suggest that the value of $\tau$ at which variability peaks is low, and so, for most of the history of a panel of decisionmakers, one should see a decline in heterogeneity. The initial rise in heterogeneity should be brief. Indeed, this is what one finds in table 4 of Bahk and Gort's paper. And Griliches and Regev find that efficiencies are

less heterogeneous among older firms.[12] Tables A-2 and A-3 in this paper also offer evidence in favor of a rise followed by a decline in the variance, although the coefficient of variation—the empirical counterpart of $\sqrt{v(N)}$—is essentially declining, in both cases.

As an explanation of cross-sectional variations in efficiency, this sort of learning heterogeneity can be quantitatively quite important. For instance, it seems that learning heterogeneity can account for a big chunk of the variance of TFP among plants engaged in similar production. To us, this finding is surprising, because the thought experiment of the model in this subsection allows decisionmakers to differ very little—they all are assumed to be doing the same thing, to have started doing it at the same time, and to have the same inherent ability. When two plants in the same industry and equal in age are compared empirically, it is unlikely that they would be using the same identical methods, be producing the exact same thing, have started operating their technologies at exactly the same time, or have an equally able management and work force. And yet even if the two plants *were* alike in all these respects, this section says that they may differ a lot in their efficiencies simply because of the inherent randomness in the learning.

How much "heterogeneity" is out there to be explained? Consider the standard deviation of the logarithm of TFP—roughly equal to the coefficient of variation of this variable. A typical estimate is around 0.30, obtained by Griliches and Regev.[13] How much can our model explain on the assumption that TFP differences among firms of a given type are caused by differences in how each of those firms is using a given technology? Estimates of the implied heterogeneity among members of a cohort are reported in the last two columns of table 1. The $\sqrt{v(N)}^{max}$ column reports the square root of the estimate of $v(N)$ when we use for $v$ the expression in the middle line of equation 16 (the largest possible heterogeneity), and the $\sqrt{v(N)}^{min}$ column reports the square root of the estimate of $v(N)$ when we use for $v$ the third line of equation

12. For the eight-year Bahk-Gort (1993) sample, the adjusted $R^2$ first rises and then falls. A more appropriate measure of heterogeneity in production functions is the mean squared error of the regression, which is what the Griliches-Regev study measured. We thank Haim Regev for making available results not reported in Griliches and Regev (1995).

13. Griliches and Regev (1995, table 5) report several estimates, all in the same ballpark. The Baily and Gersbach paper in this volume concentrates on the issue of heterogeneous efficiency.

16 (the least possible heterogeneity), which would occur even if everyone knew $\theta$ perfectly, simply because of the unpredictability of $w$ and its assumed independence over agents. The explanation that we would attribute to learning, then, really should be the difference $\sqrt{v(N)^{max}} - \sqrt{v(N)^{min}}$. A glance at table 1 reveals that in several cases, the point estimates of $\sqrt{v(N)^{max}}$ and even $\sqrt{v(N)^{min}}$ in fact badly overpredict heterogeneity. That occurs primarily because these predicted values are extremely sensitive to the values that are assigned to $\sigma_w^2$ and $N$, and that these are, in a few cases, overestimated. But we shall return to this issue later.

Finally, luck of the draw in learning really matters only when there is something to be learned, and, from this viewpoint, our model leaves out an essential ingredient. How often is technology upgraded, and how often, as a result, is an element of uncertainty injected into an agent's beliefs? How frequently is a manager of a plant faced with a new technology, and how frequently is a worker assigned to a new activity? In this sense, we have here only a chapter of the story—the chapter that describes what happens after a new technology has been set up, after a worker has been given a new job description. Still, this is an essential chapter in the story of technological change.

### Skewness in the Distribution of Efficiencies

An increase in $N$ skews the distribution of efficiencies to the right. That is, a large value of $N$ endows the distribution of efficiencies with a long right tail. To simplify the argument, suppose that each agent in the group has learned $\theta$ perfectly and consequently sets $z = \theta$. Then an agent's efficiency equals

$$(17) \qquad\qquad q = \prod_{j=1}^{N} (1 - w_j^2).$$

The $w_j$s are normally distributed and independent over $j$ and over agents. When $N = 1$, $q$ is skewed to the *left* because $w$ is normally distributed with mean zero, so that $w^2$ is skewed to the *right*. But as $N$ gets large, the distribution of $q$ acquires a right skew. If $q$ were positive, one could prove this claim by taking the logarithm of both sides of equation 17 and using the central limit theorem. But because of a probability (albeit small) that $q$ is negative, we can first square both

sides of equation 17 and then apply the central limit theorem argument to conclude that the square of $q$ is log-normal. Finally, because the probability that $q$ is negative is small, we can conclude that $q$ itself is approximately log-normal.

For $N = 1$, we can calculate the distribution of $q$ analytically because $w^2$ is distributed Chi-squared with one degree of freedom. For $N > 1$, however, the exact distribution is not known, and the density is generated by simulating the vector $(w_1, \ldots, w_N)$. For each simulation run, $N$ normal variates with mean zero and variance $\sigma_w^2 = 0.05$ were drawn to compute one value of $q$. This was repeated 20,000 times. The empirical density is reported in figure 2 for three separate values on $N$. The figure does not show it, but the probability that $q$ is negative was quite negligible: 0.0004 when $N = 50$, 0.0001 when $N = 15$, and essentially zero for $N = 1$.

Clearly, as $N$ grows, the distribution of $q$ gets skewed to the right. This is important for the ''frontier production function'' literature, where the effects of mistakes are customarily assumed to skew the distribution of efficiencies to the left. One can think of the difference between the vector of $y$s and the vector of $z$s as a ''mistake,'' and interpret it exactly as Caves and Barton interpret their ''technical inefficiency'' variable $u$.[14] We then see quite clearly that, even though the underlying shocks (the $w$s) are normal, technical inefficiency resulting from such mistakes can very well have a right skew, as Caves and Barton indeed found for distributions of firms' efficiencies in some of the industries they studied.

### The Progress Ratio

Define the progress ratio to be the multiplicative gain to productivity that unlimited experience offers. It is the ratio of eventual to initial expected productivity:

$$\text{Progress ratio} = \frac{\lim_{\tau \to \infty} E_\tau(q_\tau)}{E_0(q_0)} = \frac{(1 - \sigma_w^2)^N}{(1 - \sigma_\theta^2 - \sigma_w^2)^N} \, .$$

14. Caves and Barton (1990, p. 13). Indeed they find that efficiency, as they calculate it, exhibits a right skew for a number of industries.

**Figure 2. The Limiting ($\tau = \infty$) Distribution of Efficiencies for Three Different Values of $N$**

Frequency distribution
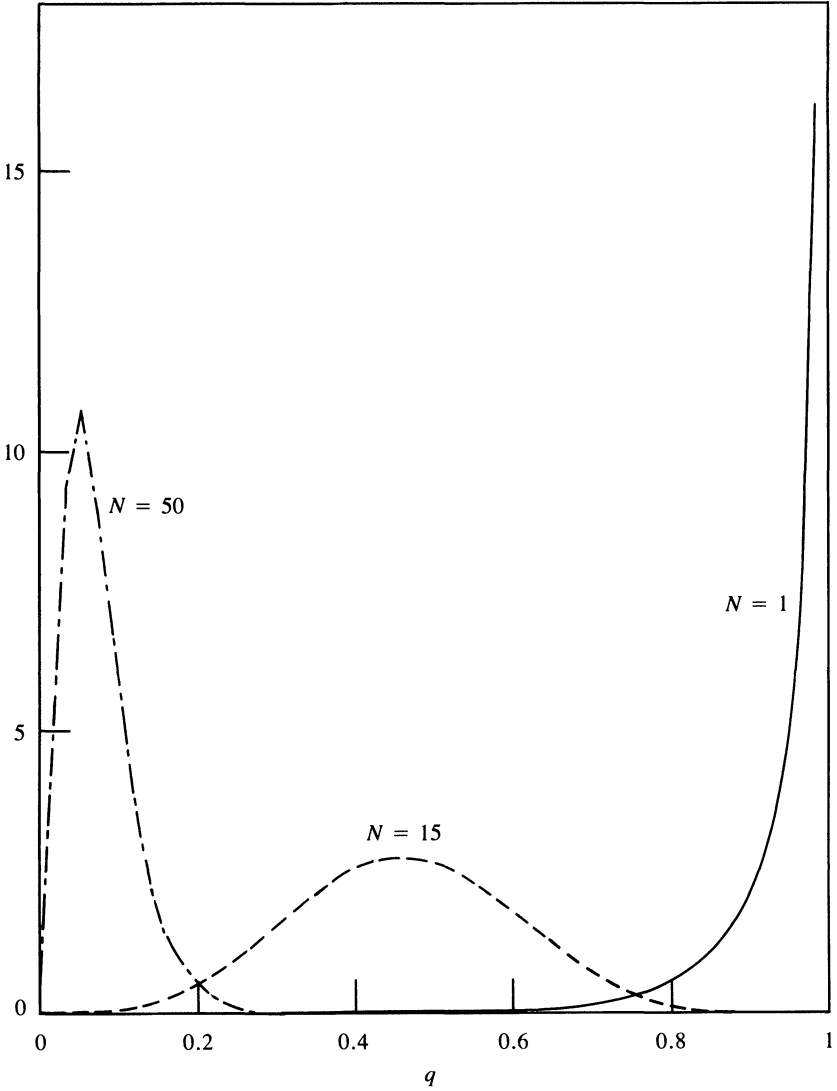


Source: Authors' calculations.

The parameter $A$ drops out from the expression for the progress ratio—
$A$ affects only the transition path from initial to terminal productivity.
The main influences on the ratio are $\sigma_\theta^2$ and $N$. If $\sigma_\theta^2 = 0$, there is
nothing to learn, and the progress ratio $= 1$. And if $N$ is large, mistakes
are compounded, and the agent must understand *all* the components of
the problem well to be efficient. Although $\sigma_w^2$ also affects the ratio, it
enters in both the numerator and the denominator, and since both it and
$\sigma_\theta^2$ are likely to be small, the progress ratio will be relatively insensitive
to $\sigma_w^2$.

Note that $\sigma_\theta^2$ and $N$ reinforce each other in this formula in the sense
that the cross partial is positive. These parameters are substitutes in
determining the progress ratio. But $N$ has an additional effect on the
shape of the learning curve—an increase in $N$ convexifies the function
$E_\tau(q_\tau)$. When $N = 1$, $E_\tau(q_\tau)$ is necessarily increasing and concave in $\tau$.
But by proposition 1, as $N$ gets large, $E_\tau(q_\tau)$ acquires an initially convex
region.

## Fitting Empirical Learning Curves as a Function of the Number of Trials

Because expected efficiency, $E_\tau(q_\tau)$ depends on $\tau$—the cumulative
number of times that the activity was undertaken—the most appropriate
data for testing the model are those that relate productivity to the cu-
mulative number of trials. When such data are available, one does not
have to infer the number of trials from observing elapsed time or cu-
mulative output. This section reports on how we fitted the model to two
such data sets.

### Learning Curves in Coronary Angioplasty

The first learning curve we fit is to "success rates" (defined precisely
in the appendix) on an operation that treats diseased blood vessels
(narrowing of the arteries). The procedure is called "percutaneous
transluminal coronary angioplasty," or PTCA, and is performed by a
doctor. The data cover about 3,100 operations performed at 105 clinical
centers by doctors to whom the technique was, at the outset, new.
Between September 1977 and September 1981, data were collected and

appendix, and the key series, the "success rate" of the operations, is shown and explained in table A-1. The total efficiency gains are modest for seemingly so complex a task. One should keep in mind, however, that the measure of "output" really is a measure of its quality only, and not its quantity. Conceivably, experienced surgeons could do the operation faster, but speed is not captured in the success measure we have.

The first two columns of table A-1 show fifteen pairs of points ($\tau$, $q'_\tau$). The $q'_\tau$ are average success rates on an operation over a large number of observations (roughly 200 operations for each $\tau$ in column 1). Data are based, then, on averaging a dichotomous variable that takes on the value 0 if the operation is a failure, and 1 if it is a success. Of course, this dichotomous variable can be thought of as taking on a value of 1 if $q$, as given by our model, is "high," and zero otherwise. Moreover, we shall assume that the mean of the dichotomous variable is proportional to the mean of $q$.

We fitted two different learning models to this curve. First, we fitted the simple multiplicative function given in equation 9, or rather its expectation $E_\tau(q_\tau)$ given in equation 12, to the points $q'_\tau$, by choosing values of the three parameters $\sigma_\theta^2$, $\sigma_w^2$, and $N$ to minimize the sum of squared residuals:

$$\sum_\tau [E_\tau(q_\tau) - q'_\tau]^2.$$

The fifteen points $\tau$ over which the summation was taken were set at the midpoints of the ranges in column 1 of table A-1.

The search was conducted over values of $\sigma_\theta^2$ and $\sigma_w^2$ between zero and 1, and for $N$ between 1 and 500. The data and the fitted $E_\tau(q_\tau)$ curve are plotted in figure 3. The minimizing values of the parameters were $\sigma_\theta^2 = 0.35$, $\sigma_w^2 = 0.27$, and $N = 1$, and the $R^2$ was 0.31. Evidently, the model does not fit well. Learning takes place too fast and ends too soon. Learning would have been slower if $N$ were larger, but in the multiplicative model, a large $N$ produces a rate of efficiency that is too low early on (see the dashed curve in figure 1 for a portrayal of what a large value of $N$ produces in the multiplicative case), and so the value $N = 1$ produces the best fit, albeit a poor one.

The efficiency function in equation 13, or rather its expected value in equation 14, fits much better: the curve portrayed in figure 4 has an

**Figure 3. The Angioplasty Data and the Line of Best Fit Based on Equation 12**

Percentage
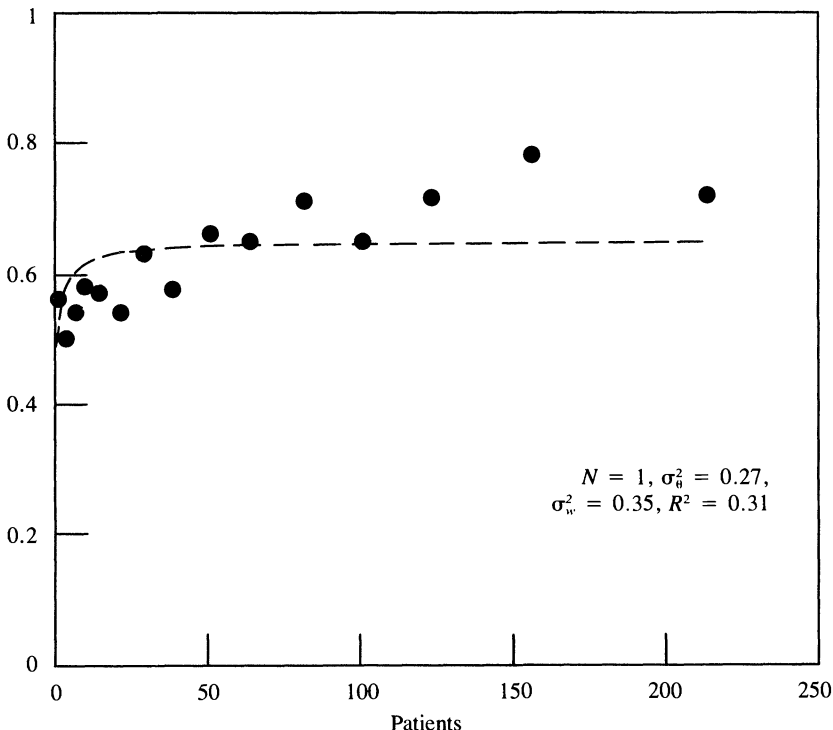


Source: Authors' calculations.

$R^2$ of 0.96. Even inexperienced doctors could manage a 50 percent success rate, and a way to get this in the model is to have productivity be the sum of a simple, easy-to-learn activity, "activity 1," that has $N = 1$, along with a second, complicated one ($N = 50$) that takes a long time to learn. The simple activity puts a "floor" on performance, while the second leads to slow, steady improvement. The other parameter values were $A_1 = 0.76$, $A_2 = 7,600,000$, $\sigma_\theta^2 = 0.6$, and $\sigma_w^2 = 0.3$.

### *Productivity and Experience in Flight Control Experiments*

The second set of data on which we actually know $\tau$ is from experiments on airplane flight control. We treat each experiment as a separate

**Figure 4. The Angioplasty Data and the Line of Best Fit Based on Equation 14**

Percentage
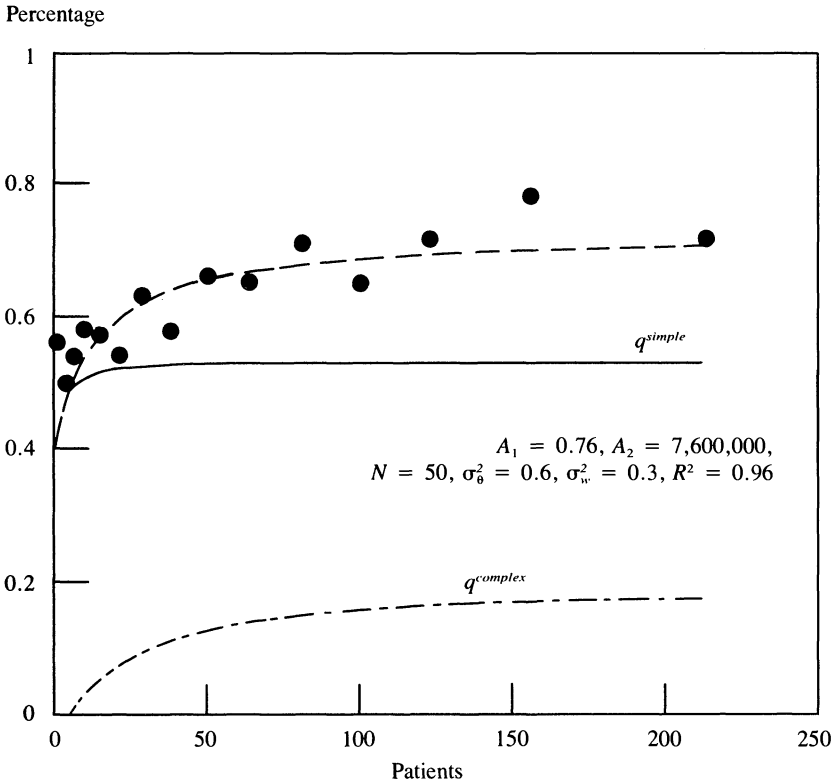


Source: Authors' calculations.

trial. Aside from having the data on a subject's cumulative number of trials (which we interpret as $\tau$), we also know that the computerized simulation of the landing of a plane could be broken down into three tasks: (a) accepting planes into the airspace, (b) moving planes in a three-level hold pattern, and (c) landing planes on the appropriate runways. Although there were three tasks, the successful simulated landing of a plane actually required at least seven decisions. (We therefore expect that $N$ should be at least 7.) Each "trial" lasted five minutes, and more than one plane could be landed during this time. The experimental subjects were 115 undergraduates, and each subject completed eighteen trials. The results of the experiments are summarized in table A-2 of the appendix. Figure 5 presents a plot of the means, along with

**Figure 5. Flight Control Experiment: Data and the Fitted Curve (Equation 12)**

Normalized number of planes landed



$N = 6$, $\sigma_\theta^2 = 0.21$, $\sigma_w^2 = 0.21$
$R^2 = 0.93101$

$\tau$ (Number of experiments)
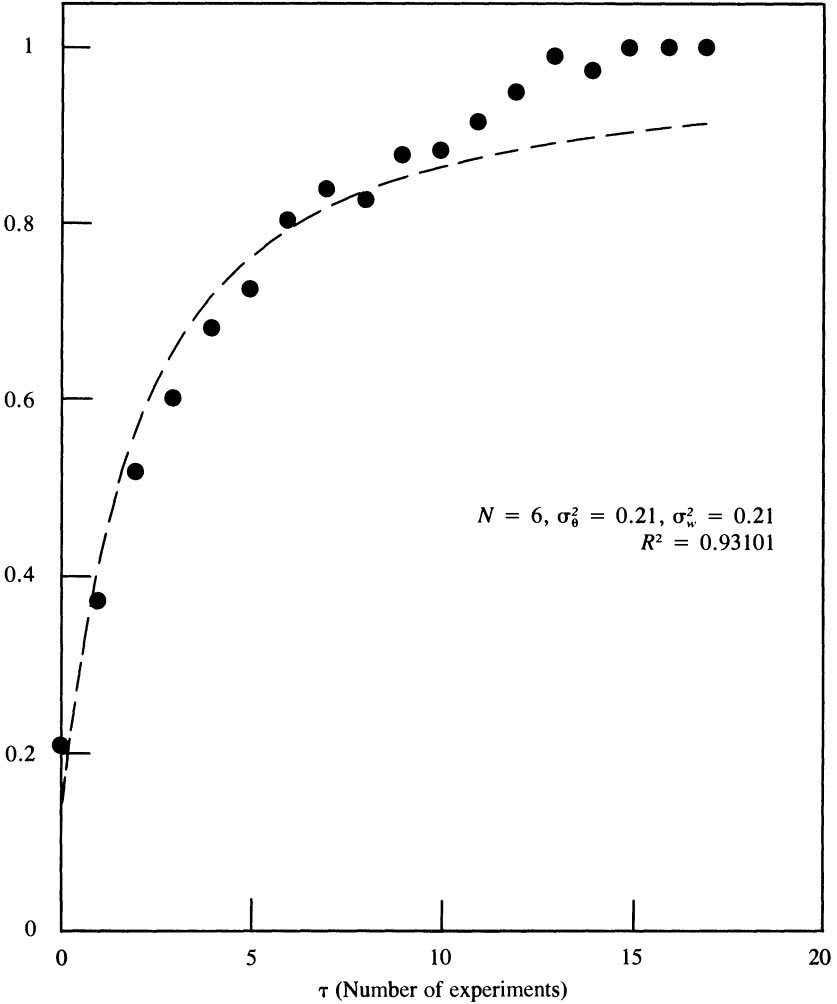
Source: Authors' calculations.

the line of best fit (again, in the least squares sense). The data were normalized by dividing each of the means by its maximal (trial 17) value of 53.13. To these data we fitted the model as expressed in equation 12. (The more elaborate model expressed in equation 14 was not needed to fit these data. Indeed, equation 12 rather than equation 14 was fitted to all of the rest of the learning curves in this paper). Because the data are normalized, we did the same thing to the expression in equation 12—we divided the expression by its maximal value, $A(1 - \sigma_w^2)^N$, and then fit the resulting curve to the normalized data. When equation 12 is normalized, the parameter A drops out, and we are left with estimates of the three other parameters in equation 12: $\sigma_\theta^2 = 0.21$, $\sigma_w^2 = 0.21$, and $N = 6$. Although the fit is good ($R^2 = 0.93$), a couple of problems remain. First, $N$ is a bit too low, although because it is not so far off its lower bound of 7, this does not seem too serious a shortcoming. Second, the model overpredicts the speed of learning during the first four or five trials and underpredicts it later on. Indeed, this problem arises with a couple of the other learning curves that we fit below. The figure shows that if we could uniformly "raise" the curve, we would get a better fit, but such a uniformly higher curve is apparently not on the menu of possibilities offered by this three-parameter family of curves.

### Learning and the Passage of Time

So far, we have measured experience only by $\tau$. Typically, however, we know only when the activity was begun and possibly about the cumulative output produced by it. The model's implications are easier stated in terms of time elapsed. This section discusses these implications.

In the model, productivity rises as the number of production runs, $\tau$, increases. The passage of time raises productivity only if it is accompanied by an increase in $\tau$. Let $\Delta$ be the length of time that it takes to complete a task. Suppose that $T$ is the number of periods of accumulated experience. Generally, the number, $\tau$, of production runs during this time interval is

$$(18) \qquad\qquad \tau = \beta T/\Delta,$$

where $\beta$ is a new parameter that is supposed to reflect the structure of production and communication among the people engaged in the activity in question. Because each task is performed once during a production run, $\tau$ is also the number of signals received about each $\theta_j$.

### Three Examples of Patterns of Specialization and Information Flows

To clarify what $\beta$ is supposed to capture, consider three very different situations, in which $\beta$ assumes different values.

- *Example 1: One worker doing the entire activity alone.* Suppose that the activity is performed by a single worker who does all N tasks in sequence. It then takes him $\Delta N$ periods to complete the activity, and if this is the only activity that this person is engaged in, $\tau = T/\Delta N$. In this example, then, $\beta = 1/N$.
- *Example 2: N workers each doing a different task simultaneously.* Because it takes only $\Delta$ periods to complete all $N$ tasks, the activity is completed once each $\Delta$ periods, and each worker performs his task once every $\Delta$ periods.[15] In this case $\beta = 1$.
- *Example 3: A worker doing the entire activity, in the company of* $m - 1$ *coworkers.* Nine out of the ten samples of workers that we analyze below fit this case: $q$ refers to the output of individual workers, but people worked alongside one another and presumably shared information. Ignore integer problems and assume that if my coworker has seen $K$ signals, and if he somehow conveys this experience to me, then this information is equivalent to my having seen $\mu K$ signals in addition to any I may have seen myself. In either event, the activity takes $\Delta N$ periods to complete, as in the first example, but because there are $m - 1$ other workers, each worker now gets $1 + \mu(m - 1)$ signals per task, at a frequency of once every $\Delta N$ periods. Then $\beta = [1 + \mu(m - 1)]/N$, so that informational spillovers speed up the learning.

Because the firm can and does decide on how to organize its production activities, it will optimally choose how complex each activity will be (how many tasks $N$ will make up each of the activities done under its roof) and the degree of specialization of its work force (whether a

---

15. To ensure that each component of the decision vector $z$ is set optimally, the firm would have to pay wages per efficiency unit of each task's services supplied.

worker ends up doing several tasks or just one of them). Hence the firm will choose $N$ and $\beta$ as a solution to some larger problem. These decisions will, however, be made before the learning processes discussed here begin, and so it is safe to treat them as predetermined, as we do here.[16]

### Deriving the Learning Curve as a Function of Time

Substituting for $\tau$ from equation 18 into equation 7 yields an expression for the posterior variance as a function of time, rather than as a function of $\tau$. Denote this posterior variance as

$$(19) \qquad\qquad x(T) = (1/\sigma_\theta^2 + \pi T)^{-1},$$

where $\pi$ measures the speed of learning:

$$(20) \qquad\qquad \pi = \frac{\dot{\beta}}{\sigma_w^2 \Delta}.$$

This is the rate at which the passage of time converts into increments of knowledge. Therefore $\Delta$, $\sigma_w^2$, and $\beta^{-1}$ affect learning in exactly the same way: a smaller $\sigma_w^2$ implies that each signal on $\theta$ is more accurate, while a smaller $\Delta$ or a bigger $\beta$ means that there are more signals per unit of time. Of course, $w$ has a direct effect on efficiency, in addition to its role of "garbling" the signal.

The derivation of $x(T)$ is based on the assumption that the rate of experimentation, that is, the number of production runs each period, is constant over time. When this assumption fails to fit the facts, learning may be better discussed in terms of the relationship between productivity and cumulative output. Although this relation is analytically more complicated in the context of our model, such a relation nevertheless exists and is discussed later in the paper. Our empirical work, however, focuses on the model's implications for the relation between productivity and the passage of time.

Let $q(T)$ denote expected efficiency after $T$ periods of experience. Then as a function of time, the learning curve is

$$(21) \qquad\qquad q(T) = A[1 - x(T) - \sigma_w^2]^N.$$

---

16. We have analyzed a model in which the degree of information sharing is determined within the system in Jovanovic and Nyarko (Forthcoming).

Now define the normalized learning curve:

$$(22) \quad q^*(T) = \frac{q(T)}{\lim_{T \to \infty} q(T)} = \frac{q(T)}{A(1 - \sigma_w^2)^N} = \left[ 1 - \frac{x(T)}{1 - \sigma_w^2} \right]^N .$$

Note that $A$ drops out of $q^*(T)$, which makes it easier to compare different learning curves. The comparative statics on $q^*(T)$ are intuitive and are summarized in the following proposition:

PROPOSITION 3: *$q^*(T)$ is decreasing in $\sigma_w^2$, $\sigma_\theta^2$, and $N$ and increasing in $\pi$.*

Care is required in interpreting the result with respect to $N$, because $N$ may also affect $\beta$, and hence $\pi$ as well. Our interpretation is: an increase in $N$ that leaves $\beta$ unchanged will raise $q^*(T)$.

## Evidence on Learning and the Passage of Time

In this section we apply the model to ten different time series of productivity on activities that were new to workers or to groups of workers doing them. The first nine pertain to activities done by workers individually, while the tenth describes the productivity of a plant, so here the decisionmaker should be considered the plant manager. The point of the exercise is not to offer new evidence on the size of the gains to learning by doing, but to highlight the flexibility of the model in fitting a diverse set of learning curves and to illustrate the implications of the model.

### The Data

The ten learning curves are presented in the form of productivity plotted against time, not the number of "experiments," as we did for the angioplasty and flight control learning curves. The first curve is for insurance sales personnel; to the extent that these agents worked alone, this curve fits the situation described in example 1. The second learning curve came from a situation that probably included elements of both examples 2 and 3: the data pertain to the total output of steel at a plant

level; no individual worker could have been responsible for all the $N$ tasks, but the workers probably shared information.

The next eight learning curves describe how individual workers' productivity in a manufacturing plant behaves as a function of their experience on a job that is new to them. Each of these eight learning curves seems to be the product of a situation that fits example 3—a worker would perform an entire activity unassisted, probably sharing some information with coworkers. One should keep the above discussion in mind when interpreting the estimates of the parameters for the individual learning curves. The data (described more fully in the appendix) were:
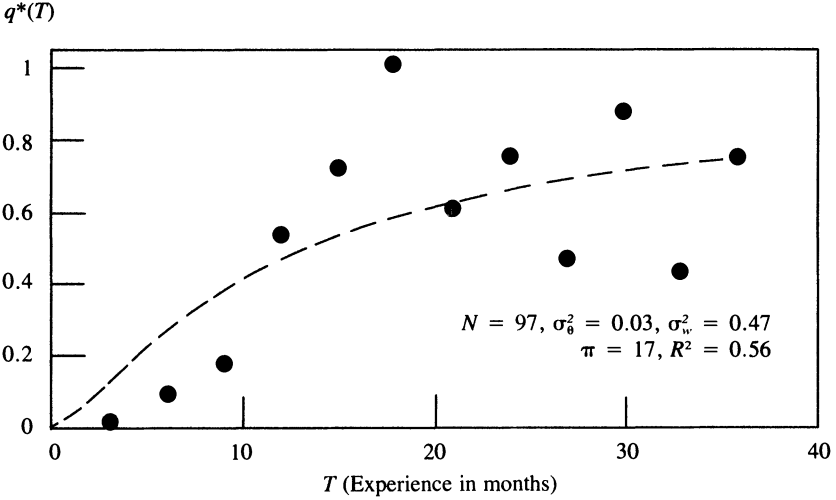
- Learning at the individual level by insurance sales personnel. Insurance sales by 319 employees of an insurance company were recorded for a period of three years and are reported, by quarter, in table A-3. The data were chronologically ordered (monthly) for each individual and then pooled into three-month averages.
- Learning at the individual level by munitions manufacturing workers. Table A-4 records the average outputs of four groups of workers in a munitions factory. The activity was new to these workers. The weekly output of each worker was recorded for nine weeks.
- Learning at the individual level by Western Electric workers. Four of the learning curves are reported in table A-5, and they record the output produced by newly hired workers in three Western Electric plants in 1977, 1979, and 1980. Each worker's output was recorded for each of the first six or seven months of his or her employment.
- Learning at the plant level in steel manufacture. Table A-6 records the month-by-month productivity of workers on a newly installed steel-finishing activity during the first two years of its operation.

### Fitting q*(T) *to the Data*

Figures 6, 7, and 8 present the plots of these data, and the lines of best fit. The fitted equation is for $q^*(T)$, given in equation 22. The estimation procedure was the same in each case and was similar to the procedure for fitting the model to the data on the flight control experiment. The only difference is that, because we do not have a measure of $\tau$ (the number of production runs), we must instead rely on the

**Figure 6. Actual and Fitted Productivity in Insurance and in the Steel Plant, by Experience, *T*, in Months**

A. Insurance sales



$N = 97, \sigma_\theta^2 = 0.03, \sigma_w^2 = 0.47$
$\pi = 17, R^2 = 0.56$

*T* (Experience in months)

B. Steel plant



$N = 43, \sigma_\theta^2 = 0.006, \sigma_w^2 = 0.77$
$\pi = 37, R^2 = 0.74$

*T* (Experience in months)
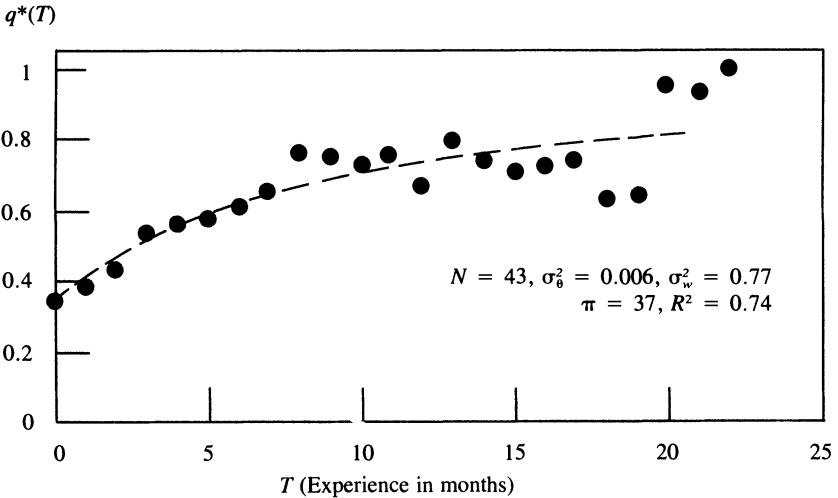
Source: Authors' calculations.

**Figure 7. Productivity of Munitions Workers, by Experience, $T$, in Months**



A. Fuse

$q*(T)$

$N = 3$,
$\sigma_\theta^2 = 0.13$, $\sigma_w^2 = 0.01$
$\pi = 91$, $R^2 = 0.98$

B. Cartridge 1

$q*(T)$

$N = 5$,
$\sigma_\theta^2 = 0.06$, $\sigma_w^2 = 0.11$
$\pi = 75$, $R^2 = 0.94$

C. Cartridge 2

$q*(T)$

$N = 7$,
$\sigma_\theta^2 = 0.04$, $\sigma_w^2 = 0.11$
$\pi = 71$, $R^2 = 0.90$

D. Cartridge 3

$q*(T)$

$N = 19$,
$\sigma_\theta^2 = 0.03$, $\sigma_w^2 = 0.13$
$\pi = 87$, $R^2 = 0.87$
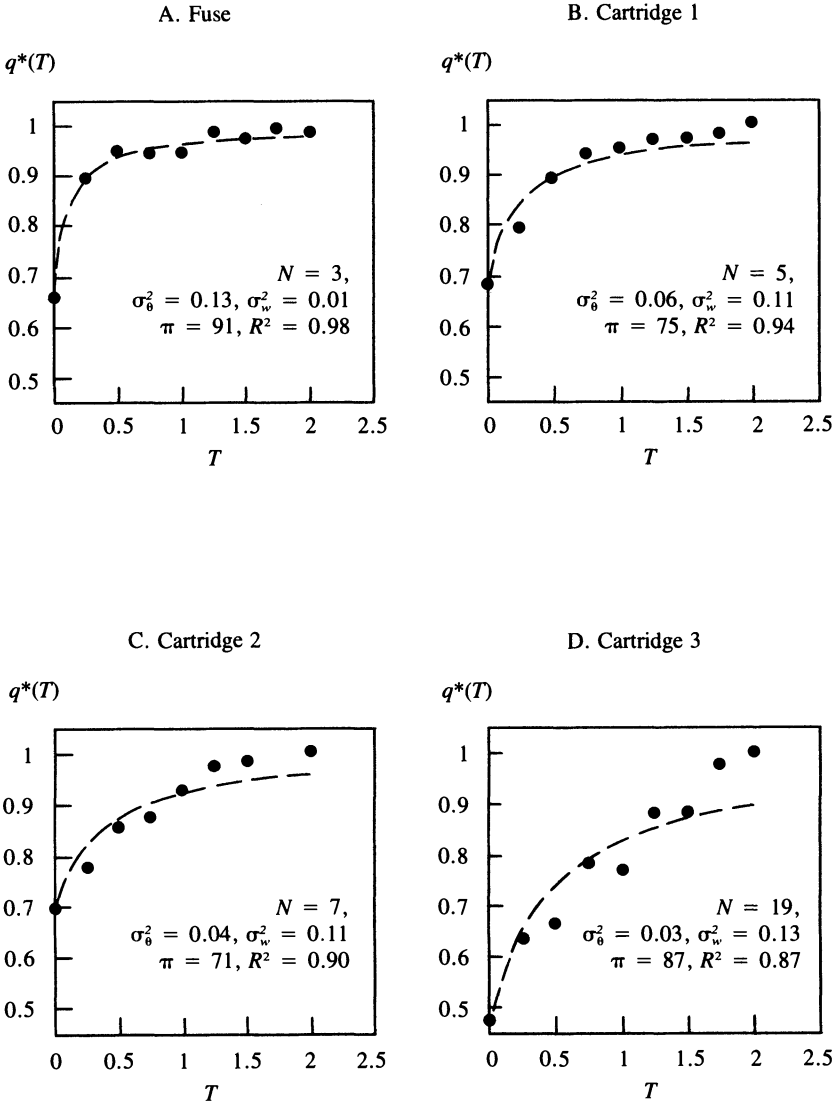
Source: Authors' calculations.

**Figure 8. Productivity of Western Electric Workers, by Experience, *T*, in Months**

A. Plant A, 1977

$q^*(T)$



$N = 1,$
$\sigma_\theta^2 = 0.39, \sigma_w^2 = 0.08$
$\pi = 55, R^2 = 0.98$

B. Plant A, 1979

$q^*(T)$



$N = 3,$
$\sigma_\theta^2 = 0.06, \sigma_w^2 = 0.01$
$\pi = 85, R^2 = 0.99$

C. Plant B

$q^*(T)$



$N = 17,$
$\sigma_\theta^2 = 0.03, \sigma_w^2 = 0.07$
$\pi = 79, R^2 = 0.96$

D. Plant C

$q^*(T)$



$N = 11,$
$\sigma_\theta^2 = 0.04, \sigma_w^2 = 0.01$
$\pi = 87, R^2 = 0.97$
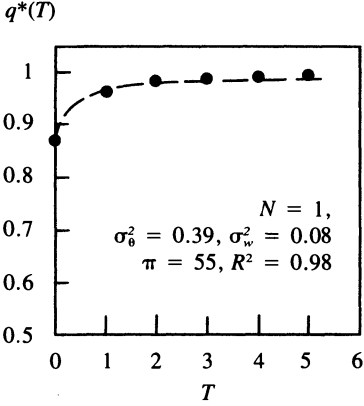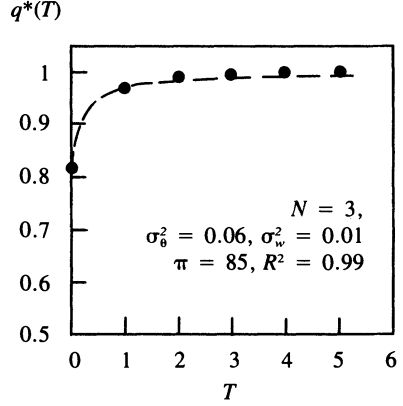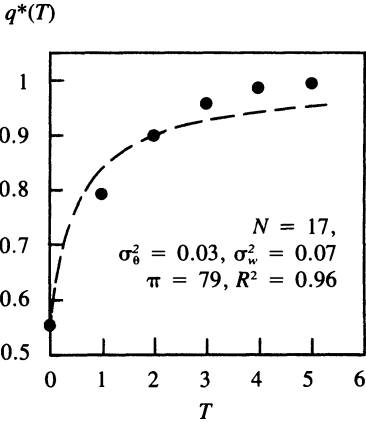
Source: Authors' calculations.

arguments for viewing learning as a function of time and use calendar time as a measure of experience. (We also could have used cumulative output instead of time, but this would have been messy). Using $T$ leads us to the version of equation 12 as expressed in equation 21 and its normalized version, $q*(T)$, in equation 22. The transformation from $\tau$ to $T$ also introduces a fourth parameter, $\pi$.

We normalized each empirical learning curve by dividing it by its highest value, which in most cases also was the terminal value. Denote this empirical learning curve by $q^+(T)$, which, like $q*(T)$, is therefore bounded above by unity. We then fitted the curve $q*(T)$ to it, by choosing $\sigma_\theta^2$, $\sigma_w^2$, $N$, and $\pi$ to minimize the sum of squared residuals:
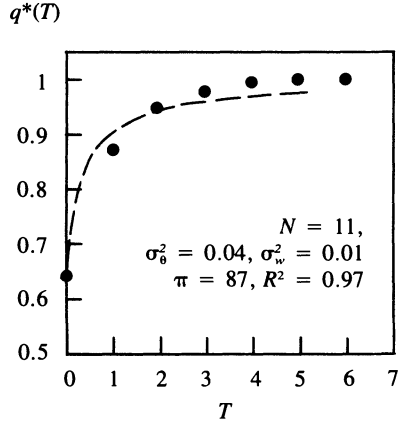
$$\text{Min}\{\underset{T}{\Sigma}[q*(T) - q^+(T)]^2\}$$

The parameters were constrained so that $q*(0) = q^+(0)$. That is, the fitted learning curves were constrained to produce the same progress ratio as the data.

Figure 6 describes the two most complex learning processes: insurance sales; and the productivity of the steel plant. Of the ten learning curves we analyze, these two had the highest progress ratios (111 in insurance sales, 2.9 in the steel plant), took much longer to learn, and consequently had the highest estimates of $N$ (97 and 43, respectively). Even though the insurance curve is an average of learning among 319 employees, it bounces around a lot, perhaps partly because of seasonality; the learning curve for the steel plant fits much better. As the surgeons' example showed (see figures 3 and 4), a large estimate of $N$ (50 for the surgeons) reflects not so much the size of the progress ratio (which for the surgeons is only 1.29), but rather the generally slow pace of learning—the half life of learning is one year in insurance and eight months in the steel plant. Moreover, there is an important difference in the types of learning in the two cases. The rising productivity of insurance agents was probably attributable largely to the rise in each agent's expertise. On the other hand, we do not know how much of the rise in the steel plant's productivity was caused by a rise in labor skills, and how much was attributable to other factors. If in equation 1, $F$ is indeed Cobb-Douglas in $K$, $L$, and $M$, then their multiplication tends to slow down the learning at the plant level for the same reason as an

increase in $N$ does. So a large estimate of $N$ for the steel plant does not necessarily mean that its workers were doing complicated things.

In comparison to insurance and steel, learning in the munitions factories and the Western Electric plants looks trivial (see figures 7 and 8). In contrast to the steel plant where we looked at the productivity per worker of the entire plant, here we actually have measures of individual output. The progress ratios are modest, and the learning takes only a few weeks (see table 1). These are simple, easy-to-learn jobs. Nevertheless, $N = 17$ for Western Electric's plant B, and $N = 19$ for one of the groups of munitions workers. Indeed, the eight fitted learning curves drawn in figures 7 and 8 do not differ that much among themselves, and yet in some cases the values of the parameters are quite different. This result suggests that not all four parameters can be precisely estimated from the $q^+(T)$ series alone, at least not when the series contains only a handful of observations. Western Electric's industrial engineers had judged the jobs done in plant C to be the most complex of the Western Electric jobs in our sample; our estimation procedure designates plant B as having had the most complex jobs.

### Interpreting the Estimates

Figures 6, 7, and 8 show that, while the fit is generally good, the model as expressed by $q^*(.)$ tends to overpredict (somewhat) the speed of learning initially and underpredict it later on. That is, the fitted line is sometimes too concave compared to the observations—the marginal returns to information diminish too fast. We showed in Proposition 1 that this feature of the model disappears as $N$ gets large; the learning curve can acquire an initial convex range, and for a while the marginal returns to information can rise. But as $N$ gets large, $q^*(0)$ tends toward zero, which limits the ability of equation 12 to fit learning patterns such as those that figures 3 and 4 show for the surgeons. At any rate, only with surgeons was this difficulty important enough to induce us to estimate the more complicated version of the model as expressed in equation 14.

We can gain further perspective if we look at all the parameter estimates together. These are gathered in table 1, which also reports some other statistics based on those estimates. The learning curves are

**Table 1. The Parameter Estimates for All Twelve Learning Curves**

| Learning curve | $N$ | $\sigma_\theta^2$ | $\sigma_w^2$ | $\pi$ | $R^2$ | Progress ratio | Half-line (in months) | $\sqrt{v(N)}^{max}$ | $\sqrt{v(N)}^{min}$ |
|---|---|---|---|---|---|---|---|---|---|
| Plant A, 1977 | 1 | 0.08 | 0.39 | 55 | 0.98 | 1.14 | 0.24 | 2.57 | 0.9 |
| Plant A, 1979 | 3 | 0.06 | 0.01 | 85 | 0.99 | 1.22 | 0.19 | 0.04 | 0.02 |
| Fuses | 3 | 0.13 | 0.01 | 91 | 0.98 | 1.51 | 0.10 | 0.05 | 0.02 |
| Cartridge 1 | 5 | 0.06 | 0.11 | 75 | 0.94 | 1.46 | 0.24 | 0.5 | 0.4 |
| Flight control[a] | 6 | 0.21 | 0.21 | b | 0.93 | 4.79 | c | 2.11 | 1.1 |
| Cartridge 2 | 7 | 0.04 | 0.11 | 71 | 0.90 | 1.43 | 0.37 | 0.59 | 0.48 |
| Plant C | 11 | 0.04 | 0.01 | 87 | 0.97 | 1.54 | 0.36 | 0.07 | 0.05 |
| Plant B | 17 | 0.03 | 0.07 | 79 | 0.96 | 1.81 | 0.52 | 0.53 | 0.46 |
| Cartridge 3 | 19 | 0.03 | 0.13 | 87 | 0.87 | 2.1 | 0.49 | 1.53 | 1.14 |
| Steel plant[d] | 43 | 0.006 | 0.77 | 37 | 0.74 | 2.9 | 8 | $7.7e^{29}$ | $2.8e^{29}$ |
| Surgeon[a] | 50 | 0.6 | 0.3 | b | 0.96 | 1.29 | c | $5.4e^{10}$ | 2494 |
| Insurance | 97 | 0.03 | 0.47 | 17 | 0.56 | 111 | 13.5 | $5.7e^{101}$ | $8e^{19}$ |

Source: Authors' calculations

[a]Learning curve estimated as a function of $\tau$, not of $T$.

[b]Parameter $\pi$ does not enter the learning curve as a function of $\tau$.

[c]Statistic can not be calculated because it is in units of $T$.

[d]Because $\sigma_w^2$ was so large, $\sqrt{v(N)}^{max}$ was given by the top line of equation 16, not the middle line.

arranged in order of increasing complexity, $N$. In the one case where a pair of learning curves has the same estimate of $N$, the curve with the larger $\sigma_\theta^2$ was assumed to be the more complex of the two. The ''progress ratio'' column reports the ratio of the maximal value of $q^+$ to its initial level, and this ratio is simply equal to $1/q^+(0)$. The ''half life'' column represents the length of time that it takes for 50 percent of the total progress to be made. Letting $T^0$ denote the half-life, $T^0$ solves the equation $q*(T^0) = [1 + q*(0)]/2$. The column labeled $\sqrt{v(N)^{max}}$ reports the point estimate of the maximal coefficient of variation of $q$ in a cohort of individuals, and the column labeled $\sqrt{v(N)^{min}}$ reports the estimate of the minimal coefficient of variation of $q$.

In table 2, we correlate the numbers reported in table 1 as we go across learning curves. The idea was to see which general features of learning were captured by which particular parameters. The surgeons' and the flight controllers' parameters were excluded from this exercise because we could not calculate $\pi$ and $T^0$ for them.

The following pattern emerges: processes on which learning is important, as measured by the progress ratio, tend to have longer half-lives of learning (bigger $T^0$), tend to be more complex (bigger $N$), and tend to give rise to more inequality in cohorts of learners (bigger $\sqrt{v(N)^{max}}$). A similar statement is true of complex processes, as measured by $N$: they have bigger progress ratios, take longer to learn, and, as we expected based on Proposition 2, produce more inequality.

Clearly, some of the numbers in the last two columns of table 1 are much too large. The last three learning curves have unacceptable second moment properties. It seems that these huge overestimates occur because the parameter $\sigma_w^2$ has been overestimated and because the $v$s are extremely sensitive to this parameter. Because the combination of estimated $(N, \sigma_w^2)$ pairs leads to such high dispersion, we looked at the possibility that the high variance estimates resulted from very large deviations of $z_j$s from the $y_j$s, which would then show up in a negative value for $1 - (y_j - z_j)^2$. Table 3 reports the probability that, evaluated at the estimated parameters, all $N$ tasks had a positive contribution (that is, that $1 - (y_j - z_j)^2 \geq 0$ for all $j$). This probability depends on $\tau$, and it is reported for $\tau = 0$, $\tau = T^0$ (the half-life of learning), and $\tau = \infty$, in the first three columns of table 3. Immediately we see that for the last three activities there is almost no chance that all the tasks will be positive at any value of $\tau$. On the other hand, the last three columns

**Table 2. Parameter Correlation Coefficients Over Ten Learning Curves**

| Parameters | $N$ | $\sigma_\theta^2$ | $\sigma_w^2$ | $\pi$ | Progress ratio | Half life | $\sqrt{v(N)}^{max}$ | $\sqrt{v(N)}^{min}$ |
|---|---|---|---|---|---|---|---|---|
| $N$ | 1 | $-0.55$ | 0.61 | $-0.83$ | 0.91 | 0.97 | 0.91 | 0.27 |
| $\sigma_\theta^2$ | $\cdots$ | 1 | $-0.51$ | 0.49 | $-0.28$ | $-0.52$ | $-0.27$ | $-0.47$ |
| $\sigma_w^2$ | $\cdots$ | $\cdots$ | 1 | $-0.86$ | 0.37 | 0.74 | 0.36 | 0.78 |
| $\pi$ | $\cdots$ | $\cdots$ | $\cdots$ | 1 | $-0.74$ | $-0.90$ | $-0.73$ | $-0.45$ |
| Progress ratio | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | 1 | 0.86 | 0.99 | $-0.09$ |
| Half life | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | 1 | 0.97 | $-0.07$ |
| $\sqrt{v(N)}^{max}$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | 1 | $-0.11$ |
| $\sqrt{v(N)}^{min}$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | 1 |

Source: Authors' calculations.

**Table 3. Probability of Positive Output on All Tasks and on a Given Task**

| | All tasks | | | A given task | | |
|---|---|---|---|---|---|---|
| Learning cure | $\tau = 0$ | $\tau = T^\circ$ | $\tau = \infty$ | $\tau = 0$ | $\tau = T^\circ$ | $\tau = \infty$ |
| Plant A, 1977 | 0.8553 | 0.8569 | 0.8907 | 0.8553 | 0.8569 | 0.8907 |
| Plant A, 1979 | 0.9995 | 0.9999 | 1.0 | 0.9998 | 0.9999 | 1.0 |
| Fuses | 0.9776 | 0.9997 | 1.0 | 0.9925 | 0.9999 | 1.0 |
| Cartridge 1 | 0.9258 | 0.9354 | 0.9872 | 0.9847 | 0.9867 | 0.9974 |
| Flight control | 0.4555 | . . . | 0.8376 | 0.8771 | . . . | 0.9709 |
| Cartridge 2 | 0.9332 | 0.9407 | 0.9821 | 0.9902 | 0.9913 | 0.9974 |
| Plant C | 0.9999 | 0.9999 | 1.0 | 0.9999 | 1.0 | 1.0 |
| Plant B | 0.9737 | 0.9807 | 0.9973 | 0.9984 | 0.9989 | 0.9998 |
| Cartridge 3 | 0.7886 | 0.8012 | 0.8997 | 0.9876 | 0.9884 | 0.9944 |
| Steel plant | $2.9527e^{-6}$ | $2.9712e^{-6}$ | $3.2846e^{-6}$ | 0.7437 | 0.7438 | 0.7455 |
| Surgeon | $3.2103e^{-8}$ | . . . | 0.0297 | 0.7081 | . . . | 0.9321 |
| Insurance | $6.1075e^{-8}$ | $1.2007e^{-7}$ | $2.6145e^{-7}$ | 0.8427 | 0.8485 | 0.8553 |

Source: Authors' calculations.

show that any given task is still likely to yield a positive contribution even for the last three activities. Still, the point is that, when two separate tasks on an activity that yield large negative outputs are multiplied together, we can get a large positive value for output. Similarly, a large negative value for a task multiplied by a positive number for the other task(s) produces a large negative realization for $q$. We are getting such unrealistically large variances because $\sigma_w^2$ and $N$ are large for the last three activities in table 2.

In response to this state of affairs, we naturally asked if the model could perform well not just in fitting the first moment of $q$—that is, the learning curve $q^*(T)$—but also its higher moments as well. For two out of the twelve learning curves—flight control and insurance sales—we also have the standard deviation of output by experience, which is displayed in tables A-2 and A-3. At the maximal level of experience (when the learning can be presumed to be mostly completed), the coefficients of variation of output were 0.23 among the flight controllers and 1.3 among the insurance agents. Some of this variance is surely attributable to differential abilities rather than to transitory variations in shocks that make up $w$. But because this is an exploratory exercise, we simply ignored the possibility that ability differs and equated these coefficients of variation to what our model would predict, which is the third line of equation 16; in other words, we equated the square root of $2\sigma_w^4/(1 - \sigma_w^2)^2$ to 0.23 for the flight controllers and to 1.3 for the

insurance agents, and we then repeated the estimation procedure using this as a constraint.

The results are shown in figure 9. Clearly, the model has no trouble meeting this constraint, even for the insurance agents, who had the largest unconstrained predicted variance. The parameter estimates change dramatically, however. Especially dramatic is the change in N for the flight controllers—from 6 to 77—and $\sigma_w^2$ drops by an order of magnitude in both cases. And yet the learning curves and the $R^2$s hardly change at all. $R^2$ falls from 0.93101 to 0.92590 for the flight controllers and from 0.56325 to 0.56321 for the insurance agents. Finally, as table 4 shows, the probability that any one task is negative is now negligible, and even the probability that all tasks are positive is respectably high.

Had second moment information been available for the other learning curves analyzed here, there is little doubt that the model could have accommodated the extra constraints with no trouble, especially because the predicted variances were quite reasonable for most of these learning curves. This outcome can be viewed as both a strength and a weakness of the model. On the positive side, this result says that a particular-shaped learning curve is consistent with varying amounts of heterogeneity in the distribution of efficiencies. But it also has a shortcoming, in that the learning curve alone will not confidently tell us all the parameters; we cannot securely tie down the parameters without having at least some information about the higher order moments in the distribution of efficiencies.

In addition to our work with this model, Foster and Rosensweig have fitted the model to the productivity of new high-yielding seed variety methods in Indian agriculture. In their version of this model, $N = 1$. Their estimates of $\sigma_\theta^2 = 0.38$ and $\sigma_w^2 = 0.11$ are comparable to some of our estimates.[17]
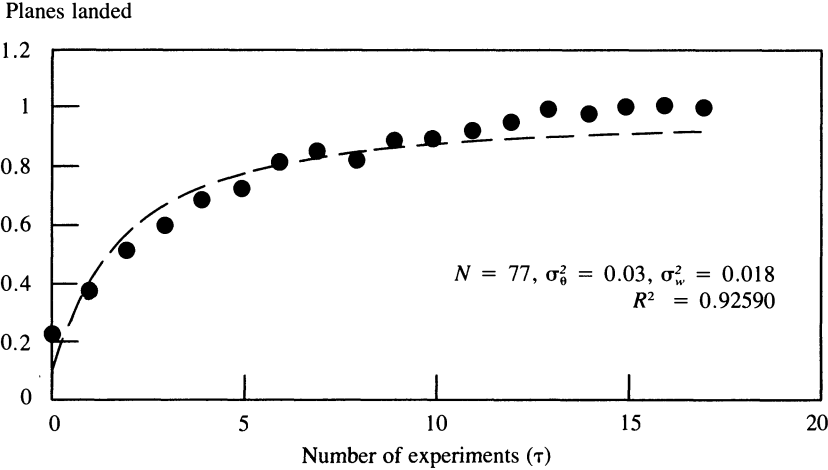
### Extensions of the Model and its Robustness

This section extends the model and elaborates on some of its properties. Each of the three subsections outlines the implications of some
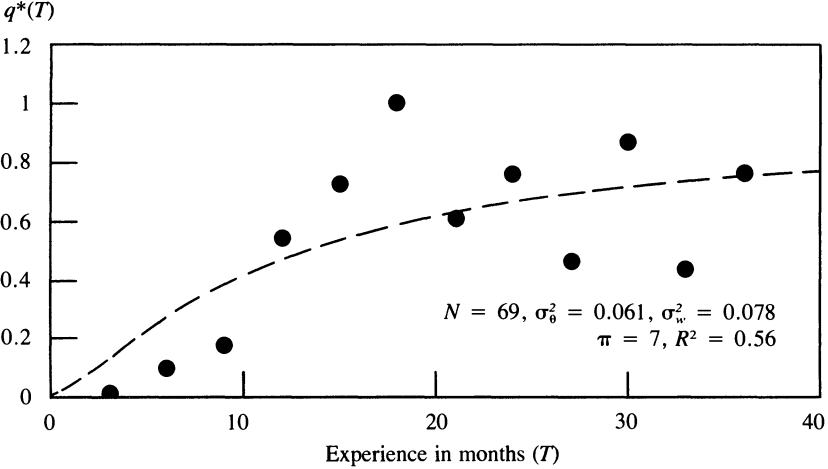
17. Foster and Rosensweig (1994) measure time in years, whereas we measure it in months, and so their estimate of $\sigma_w{}^2 = 0.091$ had to be multiplied by twelve to make it comparable to ours. We thank Foster and Rosensweig for transforming their results into the form reported here. Other related work is discussed by Zellner (1971, ch. 11).

**Figure 9. Revised Estimates for Flight Control and Insurance, Constrained by the Limiting Variance of Efficiencies**

A. Flight control

Planes landed



$N = 77, \sigma_\theta^2 = 0.03, \sigma_w^2 = 0.018$
$R^2 = 0.92590$

Number of experiments ($\tau$)

B. Insurance sales

$q^*(T)$



$N = 69, \sigma_\theta^2 = 0.061, \sigma_w^2 = 0.078$
$\pi = 7, R^2 = 0.56$

Experience in months ($T$)

Source: Authors' calculations.

**Table 4. Variance-Constrained Estimates of the Probability that Tasks Contribute a Positive Amount**

| | Probability | | | |
|---|---|---|---|---|
| | *All tasks are positive* | | *A given task is positive* | |
| *Learning curve* | $\tau = 0$ | $\tau = \infty$ | $\tau = 0$ | $\tau = \infty$ |
| Flight control | 0.9996 | 1.0 | 0.9999 | 1.0 |
| Insurance sales | 0.6026 | 0.9766 | 0.9927 | 0.9996 |

Source: Authors' calculations.

alternatives that we might have pursued, but chose not to, either because they were more complicated or because their implications were about the same as those of the model that we use.

### Learning and Cumulative Output

Instead of working with $q^*(T)$, which gives normalized productivity as a function of experience, we could derive a relation between productivity and cumulative output. Because it was more complicated and less intuitive, we did not use this approach, but it is worth outlining, especially because so much past empirical work has estimated relations between productivity and cumulative output.

Productivity rises with each production run, and so does cumulative output. Therefore the reduced form relation between productivity and cumulative output is positive. We do, however, have to be clear that we are referring to the cumulative output of the "informational unit." This may differ from the cumulative output of the plant, but that possibility is ignored here.

Let $R_\tau$ denote cumulative output of the plant over the first $\tau$ production runs, so that

$$R_\tau = \sum_{i=0}^{\tau = 1} Q_i.$$

Let $R^0(\tau) = E_0(R_\tau)$ be the cumulative output (over production runs 0 through $\tau$-1) that an outside observer would expect to see based on prior information alone, and let $Q^0(\tau) \equiv E_0(Q_\tau)$ be that outside observer's expectation about the output that will be produced at stage $\tau$. The following argument assumes that $Q = qF(K, L, M)$. A difficulty arises if $K$, $L$, and $M$ fluctuate over time and are not measured. In this case,

we could assume that they are constant, say at $K^*$, $L^*$, and $M^*$. So let $A^* \equiv AF(K^*, L^*, M^*)$. Because $E_i(Q_i)$ is a deterministic function that can be perfectly forecast at the outset, we have $E_0(Q_i) = E_i(Q_i)$ for any $i = 0, 1, 2, \ldots$, and

$$Q^0(\tau) = A^* (1 - x_\tau - \sigma_w^2)^N,$$

and

$$R^0(\tau) = A^* \sum_{i=1}^{\tau-1} (1 - x_i - \sigma_w^2)^N.$$

Treating $\tau$ as a continuous variable, we have

$$(23) \quad \frac{\partial Q^0}{\partial R^0} = \frac{\partial Q^0/\partial \tau}{\partial R^0/\partial \tau} = \frac{-\partial x_\tau/\partial \tau}{1 - x_\tau - \sigma_w^2} = \frac{x_\tau^2}{\sigma_w^2(1 - x_\tau - \sigma_w^2)} .$$

The relation between $Q^0$ and $R^0$ is not isoelastic, and, because $Q^0$ is bounded, it can be approximated by an isoelastic form, $Q^0 = \text{Const.}(R^0)^\delta$, only in the early stages.

We have chosen in the paper to work with $q^*(T)$ rather than the relation implicit in equation 23 for two reasons. First, $q^*(T)$ is available in closed form, and, second, it does not require us to assume the inputs $K$, $L$, and $M$ are constant.

### Autocorrelated w

Our initial attempt to slow down the rate of productivity growth focused on introducing autocorrelation in the signals $y$. Surprisingly, this approach did not work—the implications of the model change hardly at all. We assumed that $w_{i+1} = \rho w_i + u_{i+1}$, where the $u_i$ are mutually independent and distributed normally with mean zero and variance $\sigma_u^2$. A simple application of Bayes' rule reveals that, following $\tau$ production runs and having observed $(y_0, y_1, \ldots, y_{\tau-1})$, the decisionmaker's posterior variance over $\theta$ would be

$$x_\tau(\rho) = \left[ \frac{(1 - \rho^2)\sigma_\theta^2 + \sigma_u^2}{\sigma_\theta^2 \sigma_u^2} + \frac{(1 - \rho^2)(\tau - 1)}{\sigma_u^2} \right]^{-1} .$$

To check the plausibility of this formula, note that when $\rho = 0$, $\sigma_u^2 = \sigma_w^2$, and therefore $x_\tau(\rho) = x_\tau$. On the other hand, as $\rho \to 1$, the signals become uninformative because $\sigma_w^2 (\sigma_u^2/[1 - \rho^2]) \to \infty$, and $x_\tau(\rho) \to \sigma_\theta^2$; the posterior variance is the same as the prior variance because no

learning can take place. In any case, a higher $\rho$ means that the decisionmaker takes longer to learn $\theta$.

Despite this, the decisionmaker's productivity rises faster when $\rho$ is higher. Instead of equation 5, the optimal decision becomes

$$z_\tau(\rho) = E_\tau(y_\tau \mid \rho) = (1 - \rho) E_\tau(\theta) + \rho \, y_{\tau - 1},$$

from which it follows that

$$y_\tau(\rho) = A(1 - \{(1 - \rho) [\theta - E_\tau(\theta)] + u_\tau\}^2),$$

so that

$$E_\tau(q_\tau \mid \rho) = A [1 - (1 - \rho)^2 \, x_\tau(\rho) - \sigma_u^2].$$

This expression is almost identical to the one in equation 8, which, of course assumes that $\rho$ is zero. We conclude that the implications remain roughly intact when autocorrelation of the signals is introduced. More to the point, one *cannot* use autocorrelation of signals to get slower productivity growth—learning about $\theta$ is slower, but because past *w*s can now be used to predict future *w*s, the distance between $y$ and $z$ does not get larger, and so productivity growth is not any slower! This is why we did not take this route when trying to explain the slow pace of learning in heart surgery, and why we do not introduce $\rho$ as a parameter to be estimated.

A related alternative would be to introduce a signal on $\theta$ other than $y$ and then to vary the degree of correlation of that signal. This alternative amounts to introducing measurement error that does not enter the production function and would break the relation between the degree of autocorrelation and the ease with which disturbances to the production function could be forecast on the basis of past disturbances. The reason why this alternative does not work is that we cannot use such a signal instead of $y$. Even if we assumed that $y$ was not observed, the decisionmaker certainly observes $q$, and from $q$ and $z$, he can recover the absolute value of $y$. So all that this alternative does is add a second signal while removing a little bit of information from the first signal (that is, $y$), and so, aside from making matters analytically intractable, it would likely speed up learning, not slow it down.

### Other Loss Functions

The implications for the learning curve do not change much if, instead of being symmetric and quadratic, $q$ were given by $A(1 - a|y -$

$z|)$ if $y \geq z$, and $A(1 - b|y - z|)$ if $y \leq z$, where $y$ is still specified as
in equation 4. The optimal value of $z$ is now no longer given by equation
5, but instead equals $z_\tau = E_\tau(\theta) + (x_\tau + \sigma_w^2)^{1/2} \phi^{-1}[a/(a + b)]$, where
$\phi$ is the standard normal cumulative distribution function with density
$\varphi$.[18] Expected efficiency as a function of experience is now

(24)   $E_\tau(q_\tau) = A[1 - c(x_\tau + \sigma_w^2)]$, where $c = \varphi\{\phi^{-1}[a/(a + b)]\}$.

This expression is almost identical to that in equation 8, the only dif-
ference being that the constant $c$ is not in general equal to 1. When the
loss is linear but symmetric—that is, $a = b$—even the optimal decision
rule becomes the same as in the quadratic case: because the standard
normal distribution is centered at zero, $\phi^{-1}(1/2) = 0$, and so, when the
loss function is symmetric, $a = b$, and $z_\tau = E_\tau(\theta)$.

The presence of the constant $c$ creates a slight difference between
the learning curve in equation 8 and the learning curve in equation 24.
Even this difference can be eliminated, however, if, in the quadratic
case, in which losses rise one-for-one with $(y - z)^2$, we are free to
choose the units of $y$ and $z$ arbitrarily. (This step is justifiable if $y$ and
$z$ are not observed by the analyst). The presence of $c$ is irrelevant
because we could get a learning curve like equation 24 even in the
quadratic case by defining efficiency to equal $q = A[1 - c(y - z)^2]$.
But

(25)   $q = A[1 - c(y - z)^2] = A[1 - (y\sqrt{c} - z\sqrt{c})^2]$

$$= A[1 - (y' - z')^2],$$

where $y' = y\sqrt{c}$, and $z' = z\sqrt{c}$. The seemingly more general functional
form on the left hand side of equation 25 is therefore equivalent to the
one in equation 3, if we can rescale the units in which $z$ and $y$ are
measured. When $z$ is observed, as in Holt and others, who also use a
quadratic loss function as one component of the production function,
then one cannot choose units arbitrarily; one should instead assign a
penalty parameter such as $c$.[19] (But when neither $y$ nor $z$ is observed,
we have an identification problem in that we are unable to tell whether
big effects on productivity of learning by doing are caused by costly

18. The rest of this subsection is based on Christoffersen and Diebold (1994, sec.
3b). Manski (1988, ch. 4) discusses other loss functions and gives further references.
19. Holt and others (1960).

mistakes, that is, a large value of $c$, or whether the effects stem from the resolution of high initial ignorance about $y$).

We conclude, then, that neither the assumption of symmetry nor that of quadratic losses from mistakes is of overriding importance to the implied shape of the learning curve. Put differently, the quadratic model is more flexible than it may at first appear.

## Conclusions

In this paper we have accomplished the following four things. First, we have proposed a model of learning that, despite its simplicity, can describe a variety of activities. Second, we found that processes on which learning is important, as measured by the progress ratio, tend to take longer to learn, tend to be more complex, and tend to give rise to more inequality in cohorts of learners. A similar claim holds for complex processes: they have bigger progress ratios, take longer to learn, and generate more inequality. Third, we found that luck in learning can produce a surprising degree of inequality in efficiencies. And finally, in contrast to the assumptions of the frontier production function literature, we found that, if an activity is complex enough, the distribution of efficiencies among the decisionmakers engaged in that activity becomes skewed to the right, as it seems to be in some industries.

## Appendix: A Description of the Data

This appendix describes and presents all the data that we used in this paper. All the data were obtained from the indicated published sources.

### *Coronary Angioplasty*

The PTCA data are quite useful for our purposes.[20] First, efficiency is reliably measured. The operation aims to clear and expand a blood vessel that has, because of sedimentation or clotting, become too narrow. The operation is called a failure if, during the patient's subsequent

---

20. The data are from Kelsey and others (1984). See also the companion piece by Meier and Gruentzig (1984).

**Table A-1. The Learning Curve in Coronary Angioplasty**

| Surgeon's experience level $\tau$ | Success rate $q'_\tau$ | Number of patients in sample |
|---|---|---|
| 1–2 | 0.56 | 200 |
| 3–5 | 0.50 | 247 |
| 6–8 | 0.54 | 203 |
| 9–12 | 0.58 | 225 |
| 13–18 | 0.57 | 225 |
| 19–25 | 0.54 | 208 |
| 26–34 | 0.63 | 217 |
| 35–44 | 0.58 | 209 |
| 45–56 | 0.66 | 210 |
| 57–73 | 0.65 | 205 |
| 74–91 | 0.71 | 203 |
| 92–112 | 0.65 | 208 |
| 113–135 | 0.72 | 203 |
| 136–178 | 0.78 | 201 |
| 179–250 | 0.72 | 138 |

Source: Kelsey and others (1984).

hospitalization, a heart-bypass graft becomes necessary, or if certain other complications (including death!) occur. Otherwise, the operation is called a success. One component of efficiency that is missing, however, is the rate at which the operations were performed. In that sense, success measures the quality of a surgeon's output, but not its quantity.

Second, we know exactly how many prior operations each doctor performed—$\tau$. Because more doctors performed the operation a few times than performed it many times, sample size decreased with experience. The data were therefore pooled over successively greater and greater experience ranges, so that each point would represent roughly the same number of operations, and thus, in a sense, carry equal weight. Column 2 of table A-1 shows the fraction of successful operations by experience.

Could the rise in productivity be the result of pure selection—surgeons that reveal themselves to be adept at doing the operation could be getting assigned to more cases—and hence could this rise in the success rate simply reflect a relation between doctor's quality and the cumulative number of operations performed? But according to Kelsey and others, cases did not appear to be assigned to doctors in any systematic pattern. Similarly, some of the rise in success could have been

**Table A-2. Flight Controllers: Means, Standard Deviations, and Coefficient of Variation**

| Trial | Mean | Standard deviation | Coefficient of variation |
|-------|------|--------------------|--------------------------|
| 1 | 11.09 | 6.84 | 0.62 |
| 2 | 19.78 | 9.99 | 0.51 |
| 3 | 27.45 | 9.85 | 0.36 |
| 4 | 31.92 | 11.21 | 0.35 |
| 5 | 36.17 | 11.00 | 0.30 |
| 6 | 38.49 | 11.92 | 0.31 |
| 7 | 42.65 | 10.71 | 0.25 |
| 8 | 44.49 | 11.71 | 0.26 |
| 9 | 43.90 | 10.75 | 0.24 |
| 10 | 46.62 | 11.93 | 0.26 |
| 11 | 46.82 | 10.93 | 0.23 |
| 12 | 48.58 | 11.98 | 0.25 |
| 13 | 50.40 | 11.12 | 0.22 |
| 14 | 52.53 | 10.94 | 0.21 |
| 15 | 51.70 | 10.67 | 0.21 |
| 16 | 53.09 | 10.92 | 0.21 |
| 17 | 53.13 | 10.53 | 0.20 |
| 18 | 53.11 | 12.04 | 0.23 |

Source: Eyring, Johnson, and Francis (1993, table 1).

caused by technological advances that accompanied the rise in surgeons' experience levels, such as improvements in the equipment used during the operation (the development of a smaller catheter in 1981, for example, allowed doctors to reach more vessels, and the introduction of higher-resolution TVs made it easier to track the position of the catheter). Still, the surgeon's experience was the most important correlate of the rate of success, and for our mainly illustrative purposes, we shall treat experience as the only cause of the productivity growth that table A-1 portrays.

### The Flight Control Experiments

The results are summarized in table A-2.[21] The second column in the table shows the mean number of planes "landed" during each trial. Note that the behavior of the between-subject standard deviation is nonmonotone: as predicted by the model and discussed earlier, the

21. The data are presented and discussed in Eyring, Johnson, and Francis (1993).

Table A-3. Insurance Sales Personnel: Means, Standard Deviations, and Coefficient of Variation

| Quarter | Mean | Standard deviation | Coefficient of variation |
|---|---|---|---|
| 1 | 1,688 | 9,888 | 5.9 |
| 2 | 16,870 | 58,786 | 3.5 |
| 3 | 32,152 | 65,529 | 2.0 |
| 4 | 100,531 | 95,125 | 0.9 |
| 5 | 135,830 | 96,073 | 0.7 |
| 6 | 187,461 | 133,218 | 0.7 |
| 7 | 115,080 | 112,797 | 1.0 |
| 8 | 142,477 | 116,716 | 0.8 |
| 9 | 88,047 | 94,022 | 1.1 |
| 10 | 163,879 | 166,169 | 1.0 |
| 11 | 81,820 | 109,291 | 1.3 |
| 12 | 141,488 | 184,477 | 1.3 |

Source: Hofmann, Jacobs, and Baratta (1993, table 1).

standard deviation in the cohort first rises and then falls (before inexplicably rising again at trial 18).

### Insurance Sales

Table A-3 presents the data which were chronologically ordered (monthly) for each individual and then pooled into three-month averages, so that they are reported quarterly.[22] Although most of the people in the sample were hired about the same time (the latter half of 1984), their experience levels are not exactly synchronous. Nevertheless the data seem to exhibit some seasonality—even-numbered quarters tend to have higher sales than odd-numbered quarters.

Again we see a nonmonotonicity, as predicted by the model. The standard deviations rise steadily until the sixth quarter of experience, after which they decline substantially, and then bounce around in the last four quarters.

### Munitions Manufacture

Table A-4 records the average outputs of several groups of workers engaged in fuses and cartridges operations in a British munitions factory

22. These data are from Hofmann, Jacobs, and Baratta (1993).

**Table A-4. Weekly Output of Newly Hired Munitions Workers**

| | Relative hourly output | | | |
|---|---|---|---|---|
| *Week* | *Turning fuse bodies* | *Casings, second draw* | *Casings, second cutoff* | *Casings, reaming* |
| 1 | 67 | 70 | 71 | 48 |
| 2 | 91 | 81 | 79 | 64 |
| 3 | 96 | 91 | 87 | 67 |
| 4 | 96 | 96 | 89 | 79 |
| 5 | 96 | 97 | 94 | 78 |
| 6 | 100 | 99 | 99 | 89 |
| 7 | 99 | 99 | 100 | 89 |
| 8 | 101 | 100 | 98 | 99 |
| 9 | 100 | 102 | 102 | 101 |
| Number of workers | 16 | 12 | 14 | 16 |

Source: Vernon (1921, p. 35).

during World War I.[23] Outputs are given relative to their average during the last four weeks (or two weeks in the reaming activity). We know little about the nature of these activities except for the last one—an activity that Vernon describes as involving considerable quickness and dexterity.

### Learning in Three Western Electric Plants

Table A-5 presents four learning curves of newly hired workers in three Western Electric plants in 1977, 1979, and 1980.[24] Each worker's

**Table A-5. Median Percentage Change in Productivity of Newly Hired Workers in Three Western Electric Plants**

| *Job tenure, in months* | *Plant A, 1977 (585 workers)* | *Plant A, 1979 (308 workers)* | *Plant B (182 workers)* | *Plant C (178 workers)* |
|---|---|---|---|---|
| 1–2 | 10.8 | 18.5 | 42.5 | 35.4 |
| 2–3 | 2.0 | 2.0 | 13.0 | 8.3 |
| 3–4 | 0.8 | 0.8 | 6.4 | 3.2 |
| 4–5 | 0.4 | 0.4 | 3.0 | 1.9 |
| 5–6 | 0.0 | 0.4 | 1.0 | 0.3 |
| 6–7 | — | — | — | 0.1 |
| Progress ratio | 1.14 | 1.23 | 1.82 | 1.55 |

Source: Weiss (1994, table 5.2).

23. These data are from Vernon (1921).
24. These data are from Weiss (1994).

**Table A-6. Plant Productivity on a New Steel Finishing Activity, by Month**

| Month | Productivity | Month | Productivity | Month | Productivity |
|-------|-------------|-------|-------------|-------|-------------|
| 1 | 0.34 | 9 | 0.76 | 17 | 0.72 |
| 2 | 0.38 | 10 | 0.75 | 18 | 0.74 |
| 3 | 0.44 | 11 | 0.73 | 19 | 0.63 |
| 4 | 0.54 | 12 | 0.76 | 20 | 0.64 |
| 5 | 0.55 | 13 | 0.66 | 21 | 0.96 |
| 6 | 0.57 | 14 | 0.79 | 22 | 0.93 |
| 7 | 0.61 | 15 | 0.75 | 23 | 1.00 |
| 8 | 0.65 | 16 | 0.71 | | |

Source: Baloff (1970).

output was recorded for each of the first six or seven months of employment. The numbers are corrected for sample attrition—each entry represents the median growth rate of productivity among workers that were employed during both of the consecutive months in question. Without this correction, the growth of productivity would have been overstated, because sample attrition tended to be heaviest among the least productive workers. None of the plants was equipped with an assembly line technology. The jobs were not homogeneous in any of the plants, but they were mainly routine microscopic assembly in plant A, mainly the assembly of telephone handsets in plant B, and mainly the assembly of computer switches in plant C. Industrial engineers at the three plants made estimates of how long a new worker would take to learn to perform at the level of an experienced worker. This expected learning time was twelve weeks at plant A, seven weeks at plant B, and fifteen weeks at plant C.

### Steel Manufacture Using a New Finishing Process

These data record the month-by-month productivity of workers on a newly installed steel-finishing activity over the first two years of its operation. They are presented in table A-6, after being normalized by dividing by the highest productivity level in the last and most productive month.[25]

---

25. These data are from Baloff (1970).

# Comments and
# Discussion

**Comment by Griliches:** This paper does two things. First, it develops a Bayesian learning model (actually, several models), and, second, it fits such models to a variety of data on different learning experiments and experiences. The motivation for this exercise is that "learning by doing" is a major source of total factor productivity growth, and hence the suggested model may help account for the observed dispersion in both total factor productivity (TFP) levels and growth rates. In my comment, I largely question the latter connection.[1]

But before turning to that, I want to say a few words about learning models. The model used here is essentially a signal extraction model. We do not know what the right level of productivity is, and we use search effort, subject to stochastic error, to find it. But much of learning is not of that kind. In training we know what we want to achieve, but it requires teaching, practice, and talent. If there is uncertainty in this process, it is about our own ability to accomplish it.

It is not clear how much micromodels of learning really imply about observed macrobehavior. It is clear that there are diminishing returns (eventually) to any one particular learning activity; this is shown by the graphs in the paper. But in life, as one topic is learned, we shift to another. If one can shift to new tasks, if new technologies arrive to be learned, there need not be this kind of diminishing returns.

It would be interesting to connect the model in this paper to the

1. The original version of this paper presented at the December 1994 microeconomics meeting had a substantive section connecting the learning curve results to data on TFP dispersion from Baily, Hulten, and Campbell (1992).

300

literature on ''on-the-job training.'' That too is about learning, but although wages go up with experience and training, they do not ''concave'' very fast. Why not?

The larger question is whether such learning differences are a relevant component of the observed differences in TFP within and across industries. Because learning occurs largely at the individual level and because firms are aggregates of large numbers of individuals who come and go, much of the individual learning story averages out and does not show up in firm or industry data. Individual variability in learning will not be a major source of variance in TFP unless there is synchronization in the appearance of new technology and in rates of investment that implement it. But that does not happen very often, and when it does, it happens only in some parts of a plant and at different times in different plants. Because of such averaging, these effects are largely invisible.

Where does the observed dispersion in TFP come from? The standard TFP accounting assumes a common production function and is not explicit about the source of the remaining errors. Empirical studies indicate that there are individual firm ''fixed effects, which are at least quasi-permanent. They do not die out very fast. If learning were a major source of such dispersion, there should be much more convergence than we observe. In fact, we observe hardly any convergence even within relatively well-defined industries.

Most likely these ''errors'' are not errors made by the firms in maximizing or choosing the wrong technology but are our own ''errors.'' We, as observers, do not understand what is going on, and we do not have the tools to measure some of the relevant differences. The quality of workers my differ, the measures of capital may be wrong, capacity utilization may differ, production functions might differ, R&D-produced differences may not be accounted for, market-power positions might differ, and so on. This is then a very large haystack where the learning story may be only a very small needle. The needle is there, but it is not the major part of the story, and it will not be easy to find.

**General Discussion:** Most of the participants attempted to relate the paper's analytical framework and results to other research, including their own work. Henry Farber noted that he had been involved in studying another type of learning process—specifically, one in which the

market must learn about a worker's productivity in order to determine his wages. From this research, he has found that a normal learning model with fixed targets does not fit genuine wage data very well. That, he suggested, is because such processes do not actually involve fixed targets, but rather ones that move over time.

John Pencavel recalled a research project performed by one of his students twenty years earlier. The student had gathered several years worth of weekly data on output and earnings for individual workers at her father's apparel factory in Hawaii, where employees were paid on a piece-rate basis and a large proportion of the fashion designs produced were changed every few months. According to Pencavel, the variance of output across workers was high, but for a given worker a concave function seemed to describe well the relationships between output and time spent on the particular design. Using additional data on the location of workers within the factory, he and his student ultimately determined that managers tended to place older, more experienced workers next to newer ones so the former could teach the latter. Because some of the plant's designs changed frequently while others remained the same, Pencavel and his student were also able to address some interesting selection issues. They concluded that the most adaptable workers were those who were assigned to work on the new designs, while others who were not good learners continued to produce the old, unchanging ones.

Pencavel also pointed out that industrial relations literature suggests that the output of an individual worker is greatly affected by the social character of his plant. Workers are known to form informal ''cartels'' that attempt to sanction those who work too hard. He said that fitting the paper's data to individually based learning curves might be difficult without taking into account this factor.

Following up on Pencavel's point, Kathryn Shaw argued that the authors preferred individual learning curve data because they are easier to visualize, even though they do not conform very well to the theoretical model. From her work with the steel industry, she suggested that the learning process involves a series of learning curves, where new plateaus are continually reached. She said that when a new steel-making production line is opened, the firm initially attempts to simply get it up and running with a single work shift and one product. Subsequently, it tries to move to multiple shifts and, then, to multiple products. Finally,

the firm fine-tunes its workers to make them operate the production line as efficiently as possible.

Shaw also suggested that the authors gather a richer data set involving information on more than one production line at a startup facility. With such data, she said, it would be possible to estimate learning curves while controlling for the state of the business cycle (which affects the volume of production) and the experience of the workers (since they can be completely new or have been transferred from other plants).

# References

Ackermann, Philip. 1988. "Determinants of Individual Differences During Skill Acquisition: Cognitive Abilities and Informational Processing." *Journal of Experimental Psychology: General* 117 (3):288–318.

Argotte, Linda, and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science* 247 (February):920–24.

Arrow, Kenneth. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29 (June):155–73.

Baily, Martin Neil, and Hans Gersbach. 1995. "Efficiency in Manufacturing and the Nature of Competition." Brookings Papers on Economic Activity: Microeconomics: 307–358.

Baily, Martin Neil, Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity: Microeconomics* 1992: 187–267.

Bahk, Byong-Hyong, and Michael Gort. 1993. "Decomposing Learning by Doing in New Plants." *Journal of Political Economy* 101 (August):561–83.

Barlow, Richard, and Frank Proschan. 1975. *Statistical Theory of Reliability and Life Testing.* Holt, Rinehart and Winston.

Baloff, Nicholas. 1966. "Startups in Machine-Intensive Production Systems." *Journal of Industrial Engineering* 17 (January): 25–32.

———. 1970. "Startup Management." *IEEE Transactions on Engineering Management* 17 (November):132–41.

Beckmann, Martin. 1977. "Management Production Functions and the Theory of the Firm." *Journal of Economic Theory* 14 (February):1–18.

Bryan, W. L., and N. Harter. 1899. "Studies on the Telegraphic Language: The Acquisition of a Hierarchy of Habits." *Psychology Review* 6:345–75.

Caves, Richard, and David Barton. 1990. *Efficiency in U.S. Manufacturing Industries.* MIT Press.

Christoffersen, Peter, and Francis Diebold. 1994. "Optimal Prediction Under Asymmetric Loss." Technical Working Paper 167. Cambridge, Mass.: National Bureau of Economic Research.

Eyring, James, Debra Johnson, and David Francis. 1993. "A Cross-Level Units-of-Analysis Approach to Individual Differences in Skill Acquisition." *Journal of Applied Psychology* 78 (5):805–14.

Foster, Andrew, and Mark Rosensweig. 1994. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." Unpublished paper, University of Pennsylvania, Department of Economics.

Griliches, Zvi, and Haim Regev. 1995. "Firm Productivity in Israeli Industry: 1979–1988." *Journal of Econometrics* 65 (January):175–203.

Hofmann, David, Rick Jacobs, and Joseph Baratta. 1993. "Dynamic Criteria

and the Measurement of Change.'' *Journal of Applied Psychology* 78 (2):194–204.

Holt, Charles, and others 1960. *Planning Production, Inventories and the Work Force.* Prentice Hall.

Jovanovic, Boyan, and Yaw Nyarko. 1994. ''The Bayesian Foundations of Learning by Doing.'' Working Paper 4739. Cambridge, Mass.: National Bureau of Economic Research.

———. Forthcoming. ''The Transfer of Human Capital.'' *Journal of Economic Dynamics and Control.*

Kelsey, Sheryl, and others. 1984. ''Effect of Investigator Experience on Percutaneous Transluminal Coronary Angioplasty.'' *American Journal of Cardiology* 53 (June 15):56C–64C.

Klenow, Peter. 1993. ''Learning Curves and the Cyclical Behavior of Manufacturing Industries.'' Working paper. University of Chicago, Graduate School of Business. (September).

Kremer, M. 1993. ''The O-Ring Theory of Economic Development.'' *Quarterly Journal of Economics* 108 (August):551–75.

Manski, Charles. 1988. *Analog Estimation Problems in Economics.* New York: Chapman and Hall.

Maynard, Harold. 1971. *Industrial Engineering Handbook.* McGraw Hill.

Meier, Bernhard, and Andreas Gruentzig. 1984. ''Learning Curve for Percutaneous Transluminal Coronary Angioplasty.'' *American Journal of Cardiology* 53 (June 15):65C–66C.

Middleton, Kenneth. 1945. ''Wartime Productivity Changes in the Airframe Industry.'' *Monthly Labor Review* (August):215–25.

Parente, Stephen. 1994. ''Technology Adoption, Learning by Doing, and Economic Growth.'' *Journal of Economic Theory* 63 (August):346–69.

Rosen, Sherwin. 1972. ''Learning as Joint Production.'' *Quarterly Journal of Economics* 86 (August):366–82.

———. 1982. ''Authority, Control, and the Distribution of Earnings.'' *Bell Journal of Economics* 13 (Autumn):311–23.

Venezia, Itzhak. 1985. ''On the Statistical Origins of Learning Curves.'' *European Journal of Operations Research* 19:191–200.

Vernon, H. M. 1921. *Industrial Fatigue and Efficiency.* London: George Rutledge.

Weiss, Andrew. 1994. ''Productivity Changes without Formal Training.'' In Lisa Lynch, ed., *Training and the Public Sector,* 149–60. University of Chicago Press.

Wilson, Robert. 1975. ''Informational Economies of Scale.'' *Bell Journal of Economics* 6 (Spring):184–95.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics.* John Wiley and Sons.