RAND Journal of Economics
Vol. 31, No. 1, Spring 2000
pp. 180–205

The scope and organization of production: firm dynamics over the learning curve

Matthew F. Mitchell*

I introduce a Bayesian-learning model of the firm to account for a variety of empirical facts about firms. The many tasks the firm can undertake (the scope of the firm) are informationally related, so that the firm can enjoy some economies of scope from information. The model predicts changes in firm size and its comovement with firm scope that are broadly consistent with the empirical evidence. It also provides an explanation for the limits to the scope of the firm: the firm may lack information, or it may be costly to communicate the information necessary to undertake many tasks.

1. Introduction

■ It is now commonly known from panel data that firm size is highly variable, both for one firm over time and between firms at any point in time. The many empirical regularities concerning the evolution of firm size have given rise to models where firms have stochastic and persistent technological opportunities, such as Hopenhayn (1992), Jovanovic and MacDonald (1994), and Ericson and Pakes (1995). In this article I develop a model where these opportunities are the result of a specific mechanism, the firm's learning about its technology, and I explore the degree to which that learning mechanism can account for important features of firm dynamics.

The phenomenon of increased productivity (or lower costs) over the lifetime of the firm has been well documented (see, for instance, Argote and Epple (1990)). The intent of this article is to develop a model of learning at the firm level and see to what extent it can account for the key stylized features of the firm-dynamics literature.

The recent literature modelling the stochastic evolution of firms focuses only on the overall size of the firm. Since there are useful data on the scope of firms and the plants they operate, my model differentiates between a scope margin and a scale margin. Firms have access to a wide variety of tasks, of which they can choose a collection to undertake. The optimal way of operating each task depends on an unknown parameter.

This uncertainty poses some natural ways to explain the limitations to scope at any one firm. One reason a firm may choose to do some things but not others is that

---

* University of Minnesota; mitchell@econ.umn.edu.

This article is a chapter of my Ph.D. dissertation at this University of Rochester. I thank my advisors, Hugo Hopenhayn and Glenn MacDonald, for continuous advice and encouragement. I am also grateful for comments from Kyle Bagwell, Jeff Campbell, April Franco, Jeremy Greenwood, Tom Holmes, Boyan Jovanovic, Ron Jones, and an anonymous referee, as well as from numerous seminar and conference participants. Any remaining errors are my own.

it does not have enough knowledge to efficiently undertake tasks that are intensive in the area where its information is incomplete. In the model, some tasks are more information-intensive, in the sense that any uncertainty about the unknown parameter is magnified. Consequently, the firm is limited in its ability to incorporate progressively more tasks in its operation.

The tasks are informationally related, so that knowledge about one task is portable, to an extent, to the operation of other tasks. The firm chooses which tasks to undertake at a point in time, learns from that experience, and then chooses a new set of tasks to operate the following period. Consequently, the scope of the firm changes over time. Firms can also choose the scale at which tasks are operated. As their expertise increases on any one task, the scale of that task can rise with it. The model predicts relationships between the age of the firm, its size, and its scope that are broadly consistent with evidence from such studies as Gort (1962) and Salter and Ravenscraft (1979). It simultaneously accounts for some of the salient features of the recent panel-study evidence on the evolution of firm size. In particular, it accounts for the two most important facts about the firm size-growth relationship: smaller firms grow faster than large firms, and smaller firms also have more variable growth rates. Evans (1987) and Hall (1987) show that these facts are robust even for sample selection and measurement error correction.

The model's ability to simultaneously account for a variety of features of the data is an important part of the article. In building a model of firm scope and scale, many things will be modelled more explicitly than they have been in previous work (in particular the learning problem, and learning across a variety of tasks), but, of course, many other features of the model are rather reduced form. The model's ability to match the empirical facts, then, suggests that the learning mechanism is important and deserves further attention. It also points to future work in that, to the extent that parts of the model are less explicit, further understanding of these elements can further contribute to an empirically meaningful model of the organization of production.

As mentioned, learning and expertise are the central theme of the modelling experiment. Learning has been suggested as a force behind the stochastic nature of the fates of firms. Here the firm is modelled as a Bayesian updating profit maximizer with incomplete information about the technology it operates. In that regard, it is similar to the Bayesian-learning model of Jovanovic (1982). In that model, however, the firm's productivity is exogenously fixed; here, the firm learns how to organize production more efficiently over time. It is shown that the model presented here, aside from providing a rationale for firm growth that differs than Jovanovic (1982), can also generate some additional predictions that are consistent with firm data. Further, the informational structure here lends itself to the study of optimal firm scope. While the model focuses on the firm's accumulation of expertise on one technology, the firm is allowed to choose from many potential outputs, termed tasks, to produce at each point in time. Consequently, the limited knowledge of *how* to do tasks provides a limitation on the scope of the firm.

Since the learning process is motivated by incomplete information, the transition function regulating the evolution of the firm's productivity, and therefore employment, is a result of Bayes' rule, rather than some arbitrary law of motion. Thus the model can provide an explanation for a wide variety of empirical implications, such as the persistence of firm's outcomes, that must be assumed as part of the law of motion in learning treatments such as Ericson and Pakes (1995). On the other hand, the learning does not depend on the scale of operation, and is therefore very passive.

Uncertainty about the technology is modelled as an unknown ''target'' parameter. This part of the technology borrows heavily from recent work of Jovanovic and Nyarko

(1995, 1996) and from earlier work by Wilson (1975) and Prescott (1972). The problem can be viewed as one in which the firm must choose not only the amount of conventional inputs to hire, but also how to use those inputs. The choice of how to use the inputs can be viewed as a dial-setting problem. The firm chooses inputs and a dial setting. The best dial setting is unknown to the firm. The firm makes its best guess at the setting, observes the outcome, updates its beliefs about the optimal setting, and proceeds to the next period. The model is outlined in more detail in Section 2.

Section 3 investigates various forms of diminishing returns to scope. One possibility is that some tasks may be more difficult than others, in the sense of being more information intensive. The difficulty of a task is modelled as sensitivity to the dial setting. As tasks become more difficult, errors in the dial setting away from the best setting lead to a greater loss in output. One could say that surgery is more difficult than carpentry in the sense that small cutting errors made in surgery lead to more dire repercussions than similar errors might in carpentry. For any fixed and limited amount of information, this increasing difficulty of tasks can lead to a limitation in the number of tasks a firm operates. In other words, limited information in the face of growing difficulty limits the scope of the firm. The model provides for an explanation of the determination of firm scope, which is vital to understanding the determination of firm size.

The firm, then, chooses the tasks it undertakes, the scale of those tasks, and, importantly, the ''way'' it does the tasks, as indicated by its guess at the unknown parameter. In this regard, the model gives firms an important organizational role. Firms use their specialized knowledge, in the form of beliefs about the uncertainty they face, to organize production. How a firm goes about producing output, which might be called a production technique, is a unique and important characteristic in my model.

Another possible reason for a limited scope is the possibility that many different production techniques might be hard to unite in one firm. In informal language, a firm might ''lose focus'' if it does too many tasks that involve disparate production techniques. Such losses might arise from a variety of coordination difficulties in implementing a wide variety of production methods within the same economic unit. I formalize these reasons as a cost to disparate production techniques, so that the implications of the ''focus'' of a firm can be explored. Limited focus can once again provide a rationale for firms to operate a limited scope of tasks. There is a tradeoff between the loss of focus from expanding scope and the inherent gains to be had from the informational relationships between activities. The loss of focus is justified as a cost of communicating knowledge within disparate tasks in the firm.

I introduce a notion of technological distance to formalize this idea of focus. The distance between tasks depends on the production techniques the firm chooses for the tasks. When tasks continually differ in terms of their most efficient technique, and the operation of disparate techniques is costly, the firm again is limited in its scope. This is another way that the informational structure allows for an explanation of the limitations to scope that is unavailable when production simply maps conventional inputs to outputs.

Section 4 discusses the model's predictions. I show that the model can explain several features of the relationship between firm scope and scale. The learning curve in the model generates naturally the valid prediction that firm growth is negatively correlated with firm size. The model also has reasonable predictions for the evolution of average cost over the life of the firm. Moreover, the model suggests a rationale for the persistent outcomes of firms and suggests some additional predictions that could be examined in the data.

By combining the increasing difficulty of tasks introduced in Section 3 with the costs of setting varied techniques, the model is able to generate employment patterns that display variance across firms of the same age. Smaller firms not only grow faster, they grow more variably. The details are explained in Section 4.

2. The firm and its technology

■ Consider a firm faced with many possible tasks to undertake with a given technology. The firm produces output at each task it undertakes, and it must choose which tasks to operate. The tasks are denoted $s \in \mathbb{R}_+$. The firm chooses a closed set P of tasks to undertake in the current period. The set P is the scope of the firm, that is, the set of different things the firm does. Output for task s is produced from some inputs x_s. The ''intensity'' of task s is given by a constant-returns function $f_s(x_s)$; for simplicity, let $x \in \mathbb{R}$, so that the sole input can be thought of as labor and f can be written simply as $\alpha_s x_s$.[1]

Not only does the firm have to choose which tasks to operate and at what intensity, it must also figure out how to best undertake each task. Firm management has significant decisions to make beyond simply the hiring of inputs and the choice of tasks to operate. The firm chooses a *production technique* $q_s \in \mathbb{R}$ for each task. This technique might include organizational considerations such as the physical arrangement of the firm's inputs, the way to treat the inputs, or any other of the many management decisions that go on daily at a firm. For each task, the function $\eta_s(q_s, \theta + \epsilon)$ defines how the chosen technique q_s affects the productivity for task s. The parameter $\theta \in \mathbb{R}$ is unknown and specific to the firm. The firm learns about the nature of its technology as it learns θ, and it can translate this knowledge into higher productivity by improving its choice of production technique q_s. Improvements in productivity owing to better estimation of θ can be viewed as ''organizational capital,'' in the sense that better organization of inputs (choosing q_s more efficiently according to $\eta_s(q_s, \theta + \epsilon)$) is an input into producing output. This capital is firm specific rather than match specific, as in the specification of Prescott and Visscher (1980). The ϵ term reflects some normally distributed mean-zero independent disturbance with variance σ_ϵ^2. Output for task s is

$$y_s(q_s, \theta + \epsilon, x_s) = \eta_s(q_s, \theta + \epsilon) f_s(x_s)^\gamma.$$

The firm's production technique may be successful, or it may not, and that success affects the productivity of the inputs hired for that task. For instance, one might let $\eta_s = 1 - (s(\theta + \epsilon) - q)^2$. The production technique that maximizes expected productivity for task s, then, is $q_s = s\theta$. Since the function η_s depends on a single unknown parameter θ for all the tasks, learning about one task enhances the firm's knowledge about θ, which in turn helps the firm operate other tasks. In that sense, tasks are informationally related, and θ is interpreted as a technology-specific parameter. The organizational choice q_s can be interpreted, for instance, as the organization of a particular vintage of capital on the shop floor, or as a management practice that allows labor to be most efficient with the capital in place.

The choice of η_s dictates the informational structure of the model. Jovanovic (1982) studies a model where there is a fixed unit measure of identical tasks and the informational structure is $\eta_s = \xi(\theta + \epsilon)$, where the function ξ satisfies a variety of regularity conditions. I shall focus on forms of η_s such as the one in the previous paragraph, where θ affects the optimal choice of q_s and together they affect productivity.

[1] All of the results extend naturally to the case in which x is a vector of inputs.

This allows firms to improve their productivity through their lifetime and allows the information structure to potentially vary across s. Both of these facets of the information structure will have important implications, particularly in the firm's choice of scope P.

Another way to model informationally related activities would be to model each task as having its own unknown parameter drawn from some joint distribution with correlation between the unknown parameters for the various tasks. Dealing with multiple unknown parameters, however, limits the tractability of the model. Here, analytic results will relate the model to some data on firm dynamics. Using the single θ has the advantage of simplicity. The function η_s does the work of mapping the single θ into different optimal production techniques across tasks, since it depends on the task s.

While the informational structure will have some built-in explanations for the limitations to scope, it is less natural to find informational limits to scale. If the firm knows how to make a certain quantity of output with a set of tasks, inputs, and production techniques, then it is not clear, at least in this framework, what informational limits keep the firm from replicating that procedure. Consequently, something else is needed to limit the scale of the firm. The fact that the production function $f(x)$ is taken to the power $\gamma < 1$ indicates the usual sort of span-of-control limitations to scale, as in Lucas (1978), which are not elaborated here.[2] Output from the technology simply sums the output at all the tasks, and therefore output is given by

$$\int_P \eta_s(q_s, \theta + \epsilon) f_s(x_s)^\gamma \, ds.$$

Given a wage rate w, input costs for the technology are simply

$$\int_P w \cdot x_s \, ds.$$

The firm, then, chooses P and a set of inputs $\{x_s\}_{s\in P}$ for all the tasks in the set P. The firm also chooses a profile of production techniques $\{q_s\}_{s\in P}$ for the tasks it operates. Assume that, after realizing output, the firm also discovers, costlessly, the realized $\theta + \epsilon$ for that period.[3] The firm will be assumed to be a single-minded entity, so that we can avoid issues concerning the structure of management.[4] Suppose the firm has beliefs μ about θ represented by a normal distribution with mean $\overline{\mu}$ and variance σ_μ^2. The firm wishes to maximize the infinite sum of expected profits, given the beliefs μ, discounted by the factor $1/(1 + r)$, where $r > 0$ is the (fixed) interest rate. Denote the output price for the firm by p.[5]

Suppose that each firm in the industry draws its own θ from a normal distribution with known mean $\overline{\theta}$ and variance σ_0^2. Under this assumption, there is no incentive to watch the results of others: all the relevant information about the aggregate distribution

[2] There has, however, been a wide variety of articles focusing on this part of the firm's problem.

[3] Of course, if η_s is a one-to-one function in $\theta + \epsilon$, this is completely innocuous. However, forms that are not one to one will be the focus here.

[4] The explicit informational structure introduced here allows for natural extensions to allow for the firm to include many managers, each of which has his or her own beliefs and role in the decision process. Such an extension is an interesting avenue for future research.

[5] It is no loss to allow p and r to be time dependent. To concentrate on the role of the learning process in generating firm dynamics, the role of p and r changing over time will not be considered, so the firm can be viewed as being a member of an industry in a stationary equilibrium.

is given by $\overline{\theta}$ and σ_0^2. This allows the model to focus on the effects of the firm's own experience on learning.[6] The firm updates the beliefs μ optimally using Bayes' rule upon realizing $\theta + \epsilon$. Notice that since the observation is independent of any action, for instance the labor hire x_s, the learning is passive. Throughout, the firm will be taken to be the store of information. The internal structure of the information is left unmodelled, although a certain reduced-form for η_s introduced below is meant to capture the idea that information communication within the firm might be costly.

3. Diseconomies of scope: difficulty and focus

■ Whenever there are decreasing returns to scale at the task level ($\gamma < 1$), there is an economy of scope in the sense that spreading inputs over a larger set of tasks increases output, other things equal. The nature of this model is that the information structure contained in η_s can be used to introduce some balancing diseconomies of scope. Two are introduced. The first is increasing difficultly of tasks, so that a firm with limited information can only profitably enter into production of tasks that are ''simple'' enough; undertaking hard tasks is decreasingly productive, and so scope is limited. The second is the idea that it may be hard for the firm to ''focus'' on a variety of very different sorts of tasks, and therefore the firm must pick a limited set of tasks to undertake.

□ **Hierarchical difficulty of tasks.** In a model with multiple tasks, it makes sense to differentiate the tasks in some way. Why does the firm choose to do some tasks and not others? One possibility is that some tasks are more easily accomplished than others. The challenge for the firm is figuring out which production techniques to employ, given the uncertainty about the parameter θ. Consider the function $\eta_s(q_s, \theta + \epsilon) = 1 - (s(\theta + \epsilon) - q_s)^2$. For tasks with a higher index s, the $\theta + \epsilon$ term is magnified, making both the ''target'' choice of θ and the noise term larger. For any amount of knowledge, this will have the effect of increasing the expected disparity between q_s and the target technique as s increases. Tasks with higher index s are more information intensive.

The choice of inputs and techniques does not affect the realization of $\theta + \epsilon$. This is the sense in which learning is passive: no decision of the firm, for instance the labor choice or the dial setting, has any effect on the evolution of beliefs. As a result, the effect of any action is purely its effect on the current period's output, so the firm chooses the inputs, as well as the scope P, to maximize current-period expected output. For the choice of q_s, this is accomplished by setting $q_s = sE_\mu(\theta) = s\overline{\mu}$. E_μ denotes the expectation of θ with respect to the beliefs μ. Knowing the form of q_s, we can turn to the optimal (static) choice of scope and scale, knowing that for this formulation it is also dynamically optimal.

Expected output for task s, when q_s is taken to be the optimal $s\overline{\mu}$, is

$$(1 - s^2(\sigma_\mu^2 + \sigma_\epsilon^2))\alpha_s^\gamma x_s^\gamma. \tag{1}$$

As was suggested earlier, any uncertainty about the true θ, given by variance σ_μ^2, is magnified by the s^2 term.

[6] Irwin and Klenow (1994) find that observing the firm's own output is significantly more important to the learning process than is learning from others. Adding learning from others would be a meaningful endeavor, however. Foster and Rosenzweig (1995) find significant knowledge spillovers in agricultural data from India.

The dynamics of the beliefs μ that define the firm are generated by Bayes' rule, which means the model predicts the behavior of important firm variables. If we let $\alpha_s = 1$ for all s, then it is easy to characterize the solution to the firm's problem at any point in time. The firm operates $P = [0, b]$, where $b = 1/\sqrt{\sigma_\mu^2 + \sigma_\epsilon^2}$. The input choice x_s is $(p\gamma/w(1 - s^2(\sigma_\mu^2 + \sigma_\epsilon^2)))^{1/(1-\gamma)}$. Expected output for tasks s, then, is given by[7]

$$y_s = \left(\left(\frac{p\gamma}{w}\right)^{\gamma/(1-\gamma)}(1 - s^2(\sigma_\mu^2 + \sigma_\epsilon^2))\right)^{1/(1-\gamma)}.$$

The mean $\overline{\mu}$ does not affect the choices of the firm besides q_s, but σ_μ^2 is relevant to the labor and scope decision. Its evolution, if next-period belief variance is denoted $(\sigma_\mu^2)'$, is given by Bayes' rule as

$$(\sigma_\mu^2)' = \frac{\sigma_\mu^2\sigma_\epsilon^2}{\sigma_\mu^2 + \sigma_\epsilon^2}. \tag{2}$$

Notice that the variance of the firm's subjective beliefs falls through time. To be consistent with the information the firm has at the beginning of its life, let $\sigma_\mu^2 = \sigma_0^2$ in the first period of the firm's existence. Here, $\overline{\mu}$ doesn't enter, since the size of θ doesn't affect firm success. This is the opposite of Jovanovic (1982), where the size of θ is precisely the object of interest in determining firm's productivity. Later I develop a model that allows for ''selection,'' in the sense that some firms draw θ's that are inherently more productive while they maintain the learning structure of increasing productivity through improvements in their choice of q_s.

The fact that the firm operates a limited set of tasks is not restricted to the case where $\alpha_s = 1$. Notice that if either σ_ϵ^2 or σ_μ^2 is greater than zero, then the first part of (1) is negative for sufficiently large s. In other words, if the firm has meaningful incomplete information about the unknown parameter θ (so that σ_μ^2 is positive), then sufficiently information-intensive tasks cannot be accomplished with positive expected return. The firm will not undertake these difficult tasks, given their limited information. In the language of the model, we have the following.

Proposition 1. If $\eta_s = 1 - (s(\theta + \epsilon) - q_s)^2$ and either σ_ϵ^2 or σ_μ^2 is positive, then the scope of the firm, P, is bounded.

Information limits the scope of the firm. In this way, information is a limitation to scope along the lines of Lucas (1978), but it arises as a consequence of the uncertainty in the decision problem facing the firm. Firms have limited information, and some tasks are so information intensive that it is not profitable to operate them without more certainty about the appropriate techniques $\{q_s\}$ to use.

□ **Communication, technological distance, and loss of focus.** One force restricting the scope of firms is limited information. In addition, a firm must also communicate information to the workers who undertake production. This subsection considers a modification of the earlier model where a firm's limited ability to communicate its know-how, and not the inherent difficulty of the tasks, limits its ability to diversify.

[7] Total output is obtained by integrating y_s across tasks; the solution to this integration is quite complicated. Since it is easy to analyze the scale of each task and the overall scope of the firm with simple formulas that give unambiguous implications for total size, total output is not calculated except for the case where $\gamma = .5$, which is discussed in the next section.

When we combine the two limits to scope, the model is able to generate variance in the distribution of employment within a particular age cohort.

The role of ''focus.'' There are many reasons why implementing different production techniques may be costly to undertake within a single firm, and as a result the scope of the firm may be limited.[8] Often, the manager of the firm, who may be endowed with the knowledge given by beliefs μ, is not the one directly involved in producing the output. Communicating the production techniques for each task to the workers might be costly, since it might require the manager's time or the hiring of some sort of foreman. This is the sense in which costly communication can make undertaking various tasks costly. Sometimes, disparate production techniques are hard to integrate into one production process. If the production technique for one task involves toxic chemicals and that for another task produces food, the firm will need to go to special lengths to separate the two tasks.

The way to incorporate this in the model is to use q as a measure of distance or diversity in activities or techniques. Letting optimal techniques become farther apart as tasks become farther apart along the real line and adding a cost of setting diverse techniques forces the firm to ''focus'' on a subset of the available activities, despite the decreasing returns to any one task. To do tasks productively, q will need to be farther and farther apart as the distance between any two tasks increases.

The idea of a distance between two methods of production has appeared in many recent formalizations of learning. In Jovanovic and Nyarko (1996), new technologies involve learning a new θ. How close the new θ is to the old one plays an important role in the adoption of the new technology, since that ''closeness'' reflects how easily knowledge can be transferred to the new technology. In Auerswald et al. (1998), any two processes for producing output have an inherent, exogenous distance between them. In their framework, it is assumed that firms jump from the current process to a ''nearby'' process. Here, technological distance is in terms of the different production techniques $\{q_s\}$ used by the firm.

Modelling focus. Since this subsection is concentrating on the role of firm ''focus,'' it is reasonable to want to exclude the effects of increasing difficulty of tasks. Consider the form $\eta_s(q_s, \theta + \epsilon) = 1 - (s + \theta + \epsilon - q_s)^2$. As s increases, the best production technique increases. In that regard, tasks differ. However, errors in setting q_s are not magnified as tasks grow more complicated; all tasks are equally difficult. Assume, once again, that f_s is identical across tasks.

All the possible reasons for the costs of disparate production techniques will be put under the category of firm focus. Firms that focus on similar production techniques avoid costs associated with ''loss of focus'' when a wide variety of techniques are simultaneously employed by one firm. For a profile $\{q_s\}$, the costs are

$$C(\{q_s\}) = \phi \int_P f_s(x_s)^\gamma (q_s - \overline{q}(\{q_s\}))^2 \, ds,$$

where, letting $m(P)$ denote the Lebesgue measure of P,

[8] Among others, Rotemberg and Saloner (1994) explore the microfoundations of the costs of implementing a broad range of activities in an incomplete contracting framework.

$$\overline{q}(\{q_s\}) = \frac{\int_P q_s \, ds}{m(P)}.$$

Firms pay a cost for setting techniques away from the average technique $\overline{q}$. This can be viewed as the technological distance between techniques. The firm can implement $\overline{q}$ costlessly, for instance by distributing some sort of costless flyer or announcement to its employees, or because the employees are all doing the same thing and can easily see what technique to use from observing others around them. In that case, there is no distance between the techniques, so the distance cost is nothing. When the firm focuses on a production technique, it avoids the distance cost. Techniques far from the average require special implementation costs, including communication of the technique to employees and perhaps some special production costs in incorporating this unusual technique. The cost is increasing in the ''intensity'' $f_s(x_s)$ of the task, and it depends on the constant ϕ, which signifies how flexible the technology is to disparate production techniques. This cost represents losses due to what might broadly be called ''loss of focus'' as the firm expands the variety of things that it does.

To make the focus cost important, suppose that the input x_s must be equal across all tasks incorporated in P. The idea is that each ''production run'' involves a fixed, constant amount of each task in P. To expand scale, we can increase the number of runs, which increases inputs x identically across all tasks. The firm can still choose what tasks to do in each production run, as well as the overall scale of the operation. The importance of this assumption is that without it, firms could operate tasks at a small scale simply to manipulate their average task $\overline{q}$, and not for the profitability of the task itself. This leads to two problems. First, calculating the optimal q_s and x_s becomes very difficult because of the effect each has on all the other tasks. Second, it may be that the firm wishes to operate a great many tasks at a small scale so as to manipulate $\overline{q}$, but the tasks directly contribute almost nothing to firm output. For studying the scope of the firm, the inclusion of these ''nothing'' tasks seems uninteresting. Changing the definition of $\overline{q}$ above to weigh each q_s by $f_s(x_s)^\gamma$ does not solve the problem of firms operating tasks for indirect benefit. The impact, in terms of overall profits, of the change in $\overline{q}$ by adding an additional task is weighted over all the other tasks, since $\overline{q}$ enters into all of them and therefore may be large even if the change in $\overline{q}$ is itself small.

Here, the distance between different production methods, and the cost generated from the differences, depends on the disparity between *how* each technology is operated, as described by the production technique q_s. Technological distance plays a role in that different types of production, in the model, are hard to combine at one location. Unlike other notions of distance, the one I use here refers to differences in how a firm chooses to operate a technology. The distance between techniques is chosen by the firm, not assumed by the model. What is assumed by the model is the distance between potential techniques q and the interaction between q and s in the function η_s. The technique that minimizes expected differences from $s + \theta + \epsilon$ is moving farther and farther away as s increases.

Solving the firm's problem. Consider the case where $\alpha_s = 1$ for all s. Total output, net of focus costs, is

$$\int_P \eta_s(q_s, \overline{\theta} + \epsilon) f_s(x_s)^\gamma \, ds + \phi \int_P (q_s - \overline{q})^2 f_s(x_s)^\gamma \, ds. \tag{3}$$

Let X denote the total amount of inputs employed by the firm; since x_s is taken to be a constant x in this subsection, $X = \int_P x_s\, ds = x \cdot m(P)$. Factoring out the $f_s(x_s)^\gamma$ in (3) and replacing x_s with $X/[m(P)]$ yields

$$f\left(\frac{X}{m(P)}\right)^\gamma \left(\int_P \eta_s(q_s, \theta + \epsilon)\, ds - \phi \int_P (q_s - \overline{q})^2\, ds\right).$$

Taking advantage of the homogeneity of f, net output can be rewritten as

$$Y(X, P, \{q_s\}) = f(X)^\gamma \left(\frac{\int_P \eta_s(q_s, \theta + \epsilon)\, ds - \phi \int_P (q_s - \overline{q})^2\, ds}{m(P)^\gamma}\right). \tag{4}$$

Input costs are $w \cdot X$. Notice that the optimal choice of P and $\{q_s\}$ is independent of, and can be calculated separately from, the choice of X.

In general, finding the optimal $\{q_s\}$ is difficult, since it lies in an infinite-dimensional space. The following lemma shows how the quadratic distance cost allows for a simple form of the solution.

Lemma 1. Suppose $\eta_s = 1 - (s + \theta + \epsilon - q_s)^2$. Then for any P, there is an optimal technique profile of the (linear) form

$$q_s = \frac{1}{1 + \phi}(s + \overline{\mu}) + \frac{\phi}{1 + \phi}\overline{q}(\{q_s\}). \tag{5}$$

Proof. Consider the function

$$\hat{\pi}(P, \{q_s\}, t) = \int_P \eta_s(q_s, \theta + \epsilon)\, ds - \phi \int_P (q_s - t)^2\, ds. \tag{6}$$

When $t = \overline{q}(\{q_s\})$, (6) is the firm's objective in choosing $\{q_s\}$. By Fubini's theorem and the independence of the objective across s for a given t, the problem of $\max_{q_s} E_\mu(\hat{\pi}(P, \{q_s\}, t))$ can be solved by computing the solution of

$$\max_{q_s} E_\mu(\eta_s(q_s, \theta + \epsilon) - \phi(q_s - t)^2) \tag{7}$$

for each s. Standard maximization shows that the solution to (7) is of the form in (5). Consequently, for any $\{\tilde{q}_s\}$, we can construct a $\{q_s\}$ of the form in (5) with $\hat{\pi}(P, \{q_s\}, \overline{q}(\tilde{q}_s)) \geq \hat{\pi}(P, \{\tilde{q}_s\}, \overline{q}(\tilde{q}_s))$. Now consider the problem of $\max_t E_\mu(\hat{\pi}(P, \{q_s\}, t))$. Again, by standard techniques, it is easy to show that the solution is $t^* = \overline{q}(\{q_s\})$. Consequently, $\hat{\pi}(P, \{q_s\}, \overline{q}(q_s)) \geq \hat{\pi}(P, \{q_s\}, \overline{q}(\tilde{q}_s)) \geq \hat{\pi}(P, \{\tilde{q}_s\}, \overline{q}(\tilde{q}_s))$, and $\{q_s\}$ of the form in (5) dominates $\{\tilde{q}_s\}$. *Q.E.D.*

Suppose, without loss of generality, that the first task the firm incorporates is $s = 0$. Under the normalization that the firm incorporates task 0 first, it chooses $P = [0, b]$, and, using Lemma 1 to calculate $\{q_s\}$, b can be calculated as

$$b = 2\sqrt{3}\frac{\sqrt{1 + \phi}}{\sqrt{\phi}}\frac{\sqrt{1 - \gamma}}{\sqrt{3 - \gamma}}\sqrt{1 - \sigma_\mu^2 - \sigma_\epsilon^2}.$$

So the productivity term multiplying $f(X)^\gamma$ is

$$A* = 2^{2-\gamma}\sqrt{3^{1-\gamma}(1 - \sigma_\mu^2 - \sigma_\epsilon^2)^{3-\gamma}(1 + \phi)^{1-\gamma}\phi^{1-\gamma}(1 - \gamma)^{1-\gamma}(3 - \gamma)^{-(3-\gamma)}}.$$

The optimal scale of X can be calculated from first-order conditions as $(\gamma pA*/w)^{1/1-\gamma}$.

Under the functional form chosen for η_s, tasks are progressively more and more dissimilar. This is formalized by the fact that the optimal production technique $s + \overline{\mu}$ rises in s. To do progressively more tasks effectively, each new task's technique must be farther and farther away from the average. Eventually, the technique is either so far away from the average, or so far away from the ''best'' technique $s + \overline{\mu}$, that the firm's expected gain from adding that technique to its production runs becomes negative. Consequently, we have the following.

Proposition 2. Suppose $\eta_s = 1 - (s + \theta + \epsilon - q_s)^2$ and $\phi > 0$. Then the scope of the firm P is bounded.

Proof. Consider the second term in (4). To keep $\int_P \eta_s(q_s, \theta + \epsilon)ds$ from becoming negative and hence the contribution of tasks s to be profit reducing, q_s can never fall below $s + \theta - 1$. However, any profile that does not fall below $s + \theta - 1$ creates an arbitrarily large cost $\phi \int (q_s - \overline{q})^2\, ds$ for some task as $s \to \infty$. Since $\eta_s(q_s, \theta + \epsilon) \leq 1$, the cost eventually dominates, and the second term of (4) becomes negative for large s. *Q.E.D.*

Larger scope of production reduces focus. This tradeoff between informational economies of scope and diseconomies from the loss of focus leads to a limitation on the scope of the firm. It would also hold if the scale of each task could vary freely but had to be operated at some fixed overhead amount $\overline{o}$, so that each task had to be done at some minimum scale.

4. Empirical implications for firm dynamics

■ The model presented above provides some reasons for the limits on the scope of firms. The implications of the model can be elaborated and compared to some features of the data. Specifically, I consider scope, scale, and cost movements. Scale is taken to be measured as labor input; this is common in the empirical literature, and in the model there is a one-to-one mapping between labor input and expected output (or, by the unbiasedness of Bayes' rule, average output across many firms at a given state) for a given level of knowledge. Unless otherwise noted, the model is that of hierarchical difficulty of tasks. The model of focus has similar predictions. At the end of this section I consider a model with *both* increasing difficulty and cost of focus.

□ **The scope of firms.** Two important regularities concerning the scope of the firm emerge from the work of Gort (1962) and Salter and Ravenscraft (1979). First, bigger firms tend to be more diversified. Second, larger firms tend to also operate each activity at a larger scale then do smaller firms. In the model with hierarchical difficulty of tasks, this is a natural result. The updating equation (2) implies that σ_μ^2 declines as the firm gets older. Both scope $P = [0, 1/(\sigma_\mu^2 + \sigma_\epsilon^2)]$ and the firm's labor hired at task

s, $x_s = (p\gamma/w(1 - s^2(\sigma_\mu^2 + \sigma_\epsilon^2)))^{1/1-\gamma}$, rise through time as the firms variance σ_μ^2 decreases. Older firms are larger and operate more tasks, operate at a larger scale, and operate a given task at a larger scale because their knowledge, embodied in a lower σ_μ^2, allows them to do more difficult tasks efficiently. Older, knowledgeable firms are bigger and involved in more things by virtue of their accumulated experience as measured by lower variance about the unknown parameter

Another feature of the model that is consistent with both casual empiricism and some studies of particular industries is that firms commence production with a subset of the tasks they eventually undertake. For instance, Shaw (1995) suggests that new steel plants start up doing only a few of the tasks they intend to eventually do, learn from experience, and then add tasks. This is exactly the sequence of events the model predicts: not only does P rise in terms of its measure, but also $P_t \subset P_{t+1}$. Along the scope margin, the model naturally predicts that firms move from the simplest tasks to more complicated ones. A different Bayesian-learning explanation for this progression is presented in Jovanovic and Nyarko (1997). In their framework, agents start with simple activities because simple activities offer better learning possibilities. In the model above with progressively more complicated tasks, firms start with simple activities because the loss from making bad decisions on more complicated tasks is exaggerated. Firms wait until they have further knowledge before undertaking the more difficult tasks.

It should be noted that as a model of diversification, this model points to a learning mechanism driving diversification into *related* tasks as the firm accumulates information, since the relevant scope-determining parameter is knowledge about a parameter θ common to all tasks. Diversification into similar tasks appears to be a quantitatively important element of diversification. Figure 1 is a histogram of diversification data taken from a CompuStat manufacturing sample studied by Cardinal and Opler (1995). The horizontal axis is defined by the fraction of a particular firm's employment that resides in its largest (as measured by employment) four-digit Standard Industry Classification (SIC) code line of business. At this level of disaggregation, only a very small fraction of firms (the rightmost bar—about 2%) are nondiversified in this sample. In fact, many firms have less than half of their employment in any one four-digit SIC code, as seen by the bars to the left of .5. Figure 2 repeats the exercise, this time using the less aggregated two-digit SIC codes. More firms are nondiversified at the two-digit level, significantly more than if firms were choosing their four-digit product line choices randomly. This suggests that related diversification is important: The choice of industry is not random across four-digit industries but rather is clustered in two-digit industries, suggesting that the sort of learning mechanism proposed has something relevant to say about diversification.[9]

□ **Average costs and the learning curve.** Aside from the introduction of a scope margin, a difference between my model of firm learning and the most popular learning model used in the study of firm dynamics, the selection model of Jovanovic (1982), comes in the evolution of productivity (and, as a result, average costs) over the learning curve.

The selection model concerns firms learning their (exogenously fixed) productivity. As a firm learns its productivity, it learns the optimal amount of inputs to hire to minimize per-unit costs. In the formulation above, productivity is not fixed but tends to rise as firms learn how to operate their technology. In Jovanovic's selection model,

[9] Certainly, examples of unrelated diversification can be found, relying on other explanations.

FIGURE 1

DIVERSIFICATION, FOUR-DIGIT LEVEL

growth comes about through firms discovering that they have low costs, and a functional form restriction that leads to firms with lower costs operating at a larger scale.

A standard result in the study of learning curves is that average costs tend to fall over time. If one were to calculate the mean of average variable cost across a cohort of firms, it would equal any one firm's costs divided by its expected (value of) output.[10] Consequently, mean average variable costs (AVC) are

$$AVC = \frac{w \int_P x_w \, ds}{p \int_P y_s \, ds}.$$

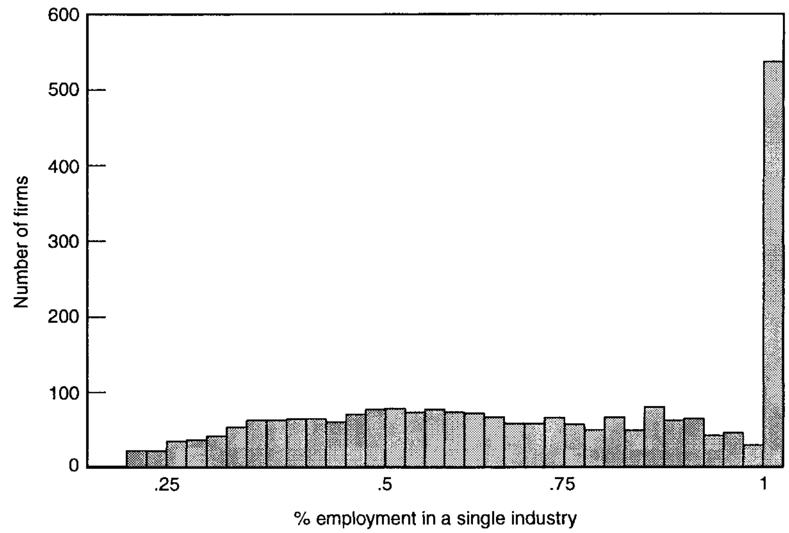
Learning curve studies report the evolution of average costs over time, evaluated at the firm's choice of inputs through time. In each case of the model, y_s reduces to $y_s = A_s x_s^\gamma$, where A_s is a constant depending on the task, the dial setting, any focus costs, and so on. The first-order condition for choice of x_s is $pA_s \gamma x_s^{\gamma-1} = w$, implying $y_s = (w/p\gamma)x_s$. Substituting this for y_s yields average variable costs at the firm's optimum, AVC_t^*:

$$AVC_t^* = \frac{w \int_P x_s \, ds}{p \int_P \left(\frac{w}{p\gamma}\right) x_s \, ds} = \gamma.$$

[10] In many cases with firm-level data, output is measured through dollar values; multiplying the results by p transforms the measures to the ones obtained using physical units of output.

FIGURE 2

DIVERSIFICATION, TWO-DIGIT LEVEL



In the model, average variable costs, at the optimum, are constantly γ over the firm's life: as the firm learns, its upward-sloping cost curve shifts down. The offsetting effects of lowering the cost curve and moving out along it as output rises leave average variable costs unchanged.[11] In the selection model, firms have a fixed, upward-sloping, unknown average variable cost curve. Survivors expand along this curve, and as a result, measured average variable costs *rise* over a firm's life.

If the firm must pay a fixed cost F in order to operate each period,[12] measured average costs are the sum of average variable costs and average fixed costs:

$$AFC = \frac{F}{p\int_P y_s \, ds}.$$

Again measured at the optimal choice, this falls over time as firms expand output in either my model or the selection model. As a result, the evolution of measured average costs over time is ambiguous in the selection model, but it falls unambiguously in my model if $F > 0$.

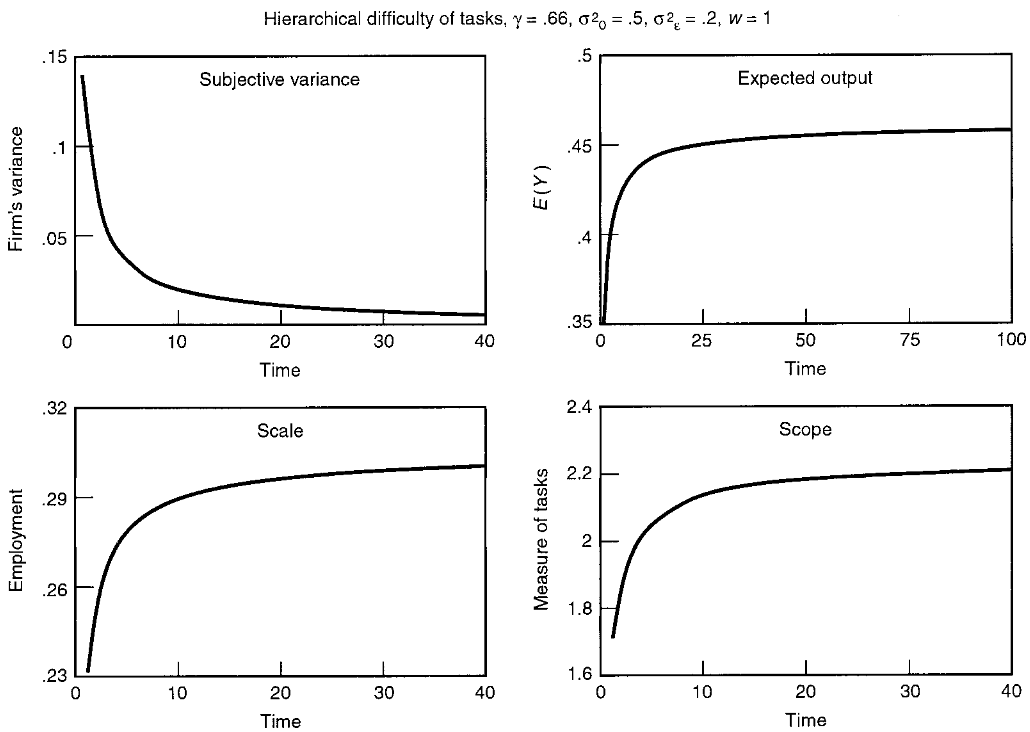
□ **Gibrat's Law.** The results for the model can also be compared to the literature on the growth of firm size. Figure 3 depicts the learning curve and its impact on scope, scale, and expected output for the firm. Because of the gradual slowdown of learning that results from the Bayes' rule updating equation for σ_μ^2, the firm grows more quickly when it is young and small. This is consistent with much of the evidence from firm

[11] This logic is equally applicable if x is a vector of inputs, so that each task is produced $A_s f(x)^\gamma$. The derivation is an application of Euler's Law for homogeneous functions.

[12] The imposition of this sort of fixed costs is natural in equilibrium models of firm dynamics, since it forces a distinction between an exiting firm (which avoids paying F) and a firm that produces nothing but stays in the industry, and hence must pay F.

FIGURE 3

FIRM DYNAMICS OVER THE LEARNING CURVE



panel data.[13] Those studies find that growth independent of size, called Gibrat's Law, is generally violated for small firms, but it is not so clearly overturned for larger firms. Figure 3 suggests that larger, older firms further along the learning curve find themselves on a plateau, which leads to low and constant growth rates.[14]

□ **The stochastic process of firm dynamics: productivity and employment.** Next let us consider the model's implications for the stochastic nature of firm outcomes. One of the most basic facts about firm dynamics, exploited in models such as Hopenhayn (1992) and Ericson and Pakes (1995), is that a firm's fates are stochastic and persistent. This is a feature of my model as well.

Productivity. Models frequently use persistent shocks at the firm level. My model has only independent shocks, but the effect of the shock is persistent. Consider a firm in the increasing-difficulty model (for now without any focus costs) with beliefs given by mean $\overline{\mu}$ and variance σ^2_{μ}, which has actually drawn the parameter θ. The firm choice is as described before, and output on task s is given by

$$(1 - (s(\theta - \overline{\mu} + \epsilon))^2)x^{\gamma}. \tag{8}$$

[13] See Evans (1987) and Hall (1987); for a more general description of the size/growth literature, Sutton (1997) reviews the Gilbert's Law literature.

[14] Jovanovic and Nyarko (1995) refine a similar Bayesian-learning problem in order to fit some empirical learning curves. Those results could be incorporated here, if the intent were to carefully match this model's predictions for firm variables to their empirical counterparts for some particular data. Instead, this article focuses on qualitative relationships between firm variables.

The formula for output (8) has a familiar interpretation: $(1 - (s(\theta - \overline{\mu} + \epsilon))^2)$ serves as a time-varying productivity term.

The firm's productivity is serially correlated despite the independent shock ϵ. The firm, observing $\theta + \epsilon$, updates its beliefs. The evolution of the mean of the beliefs, $\overline{\mu}$, is now important. The new mean $\overline{\mu}'$ is given by (from Bayes' rule)

$$\overline{\mu}' = \lambda\theta + (1 - \lambda)\overline{\mu} + \lambda\epsilon, \tag{9}$$

where $\lambda = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\epsilon^2)$. For a given θ, serial correlation of $\overline{\mu}$ leads overall firm productivity to be correlated.

Here the model is in contrast to the typical assumptions about the process that are assumed, for instance, in Hopenhayn (1992)–that the evolution of firm productivity is constant over the firm's lifetime. The correlation of $\overline{\mu}$, and hence productivity, rises over time. Eventually σ_μ^2 approaches zero and so $(1 - \lambda)$ approaches one, meaning that the long-run correlation of productivity is one. This arises since $\overline{\mu}$ converges to θ, so that eventually the problem the firm faces is constant over time. At the beginning of a firm's life, productivity is largely due to randomness (luck), since knowledge is relatively low. As the firm ages and accumulates knowledge, the effect of accumulated knowledge makes the ''luck'' component relatively less important, and hence the correlation rises.

Another new implication of the learning model for the firm-productivity process is that it leads the variance of output of a given firm to be autocorrelated. In the case where $\gamma = .5$, integration by parts yields a variance of productivity of

$$\text{Var}(Y) = \frac{2}{225}\left(\frac{p}{w}\right)^2 \sigma_\epsilon^2 \frac{2(\theta - \overline{\mu})^2 + \sigma_\epsilon^2}{(\sigma_\epsilon^2 + \sigma_\mu^2)^3}.$$

Since $\overline{\mu}$ is autocorrelated, the variance of output is autocorrelated for each firm. In the language of econometrics, output and productivity should show ARCH,[15] despite the shocks being independent. Intuitively, firms with bad estimates of θ, that is, firms that do not know very well how to operate their technology, have high variance of output. Those firms are likely to have bad estimates in the next period, and hence high variance, so the variance is autocorrelated.

Employment. When either the increasing-difficulty or the loss-of-focus models is used to explain the limits to firm scope, expected productivity increases monotonically. As a result, the model predicts that ''older'' is always synonymous with ''bigger'' (in terms of employment) and ''more productive.'' Although firms do tend to grow as they age, this monotonicity is inconsistent with data such as that presented by Hall (1987), where employment varies both up and down.

Unlike the model with only increasing difficulty of tasks, a model with both increasing difficulty and costs to loss of focus will allow the firm's estimate of the unknown parameter, $\overline{\mu}$, to influence not only realized productivity but expected productivity, and therefore affect employment decisions. Since $\overline{\mu}$ fluctuates randomly, this will give rise to randomness in employment. The Bayesian-learning structure will generate predictions for first- and second-moment properties of employment dynamics that can be compared to the stylized facts of empirical studies on firm dynamics. This is a

[15] Autoregressive conditional heteroskedasticity.

real payoff of the Bayesian structure: since variances affect mean productivity, the first- and second-moment predictions are tied together. We can use the model to help understand some of the key facts in such panel data studies as Evans (1987) and Hall (1987). Specifically, small firms grow faster than large firms on average, and small firms' growth rates are more variable than those of large firms.

Consider the multiplicative form $1 - (s(\theta + \epsilon) - q_s)^2$, combined into a model with costs of focus, i.e., where $\phi > 0$. Again, for simplicity, suppose that x_s must be constant across all tasks, i.e., $x_s = x$, for each production run. Tasks are progressively more difficult and cause a loss of focus. The firm chooses a set of tasks of the form $[0, b]$. As the firm incorporates more and more tasks, the tasks become increasingly difficult. Simultaneously, the firm encounters loss of focus, since conditional on a given expectation of θ, the firm is drawn to choose $\{q_s\}$ close to $s\theta$ in order to minimize losses from the $(s(\theta + \epsilon) - q_s)^2$ term.

Variance in beliefs about θ generates lower efficiency as before, since it leads q_s to be farther from $s\theta$ on average. In addition, however, the mean of beliefs $\overline{\mu}$ is also important. When θ is larger in absolute value, the set of processes $\{q_s\}$ that minimizes losses from the $(s(\theta + \epsilon) - q_s)^2$ term is steeper, and steeper profiles $\{q_s\}$ have a higher cost resulting from loss of focus. Firms learning that they have drawn a large θ (in absolute value) will be forced to be smaller than firms that have drawn a smaller θ. As a result, firms adjust their scope and scale in response to changes in their estimate of the mean $\overline{\mu}$ of θ as well as to the variance σ_μ^2, and so size is affected by the random fluctuations given in (9), which can go either up or down. At the same time, the fact that lower σ_μ^2 allows the firm to choose $\{q_s\}$ more precisely will still tend to increase firm size.

Once again, the quadratic form of the two loss functions from $\{q_s\}$ leads the firm to choose a linear combination of the "no focus costs" optimal choice $s\overline{\mu}$ and the average technique $\overline{q}(\{q_s\})$ defined earlier. Using the fact that the firm chooses an interval $[0, b]$, the optimal profile can be characterized as follows.

Lemma 2. Suppose $\eta_s = 1 - (s(\theta + \epsilon) - q_s)^2$. Then the optimal technique profile is

$$q_s^* = \frac{1}{1 + \phi}(s\overline{\mu}) + \frac{\phi}{(1 + \phi)}\frac{b\overline{\mu}}{2}. \tag{10}$$

Proof. A proof identical to Lemma 1, everywhere replacing $s + \overline{\mu}$ with $s\overline{\mu}$, shows that q_s is of the form

$$q_s = \frac{1}{1 + \phi}(s\overline{\mu}) + \frac{\phi}{1 + \phi}\overline{q}(\{q_s\}).$$

Since the firm chooses to produce over an interval $[0, b]$, $\overline{q}$ can be calculated by integrating q_s:

$$\overline{q} = \frac{1}{2(1 + \phi)}\overline{\mu}b + \frac{\phi}{1 + \phi}\overline{q}.$$

Solving for $\overline{q}$ gives $\overline{q} = \overline{\mu}b/2$, and the result. *Q.E.D.*

To see the effect of $\overline{\mu}$ on the firm's employment and output, consider the absolute difference between $s\overline{\mu}$ and q_s:

$$|s\overline{\mu} - q_s| = \left|\frac{\phi}{1+\phi}(s\overline{\mu}) - \frac{\phi}{(1+\phi)}\frac{b\overline{\mu}}{2}\right| = \frac{\phi}{1+\phi}|\overline{\mu}|\left|s - \frac{b}{2}\right|.$$

This is increasing in $|\mu|$: whenever $|\mu|$ is bigger, focus considerations cause the optimal profile to be farther away from $s\overline{\mu}$. On the other hand, consider the difference between q_s and the average technique $\overline{q}$:

$$\left|q_s - \frac{\overline{\mu}b}{2}\right| = \left|\frac{1}{1+\phi}(s\overline{\mu}) - \frac{\phi}{(1+\phi)}\frac{b\overline{\mu}}{2}\right| = \frac{1}{1+\phi}|\overline{\mu}|\left|s - \frac{b}{2}\right|.$$

Once again, this is increasing in $|\overline{\mu}|$.

Total output is

$$\int_P (1 - (s(\theta + \epsilon) - q_s)^2 - \phi(q_s - \overline{q})^2) f_s(x_s)^\gamma \, ds.$$

The effective productivity for each task, $(1 - (s(\theta + \epsilon) - q_s)^2 - \phi(q_s - \overline{q})^2$, is decreasing in $|\overline{\mu}|$, and therefore both x_s and b are decreasing in $|\overline{\mu}|$ as well. As in the selection model, some firms draw a more favorable θ than others. A larger draw of θ indicates a less-productive technology, since the organizational structure q_s needed to operate varied tasks efficiently involves very disparate, and thus costly, production techniques.

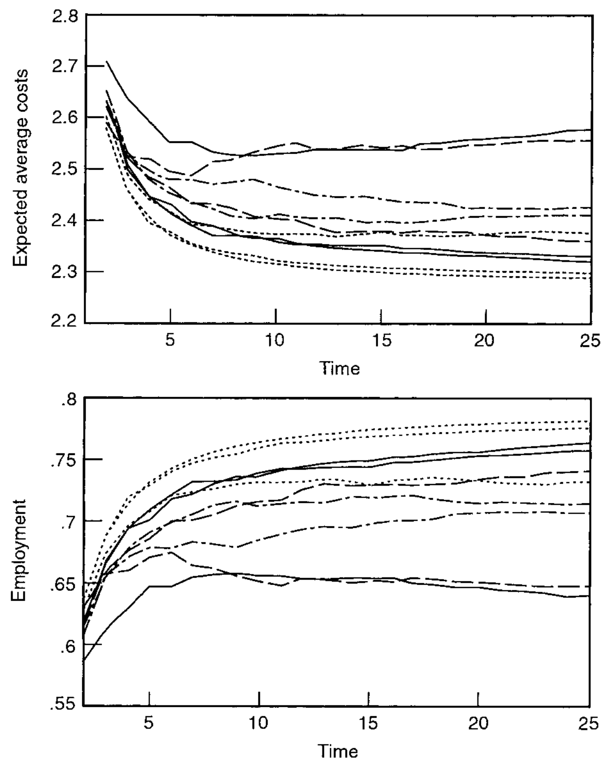
Since $\overline{\mu}$ affects the firm's scope and employment choices, all the stochastic results deriving from the nature of the evolution of $\overline{\mu}$ then pertain to both the scope of the firm and employment at the firm. Since $\overline{\mu}$ is stochastic and varies in both directions, it is possible to explain stochastic employment moving both up and down, as in the data. Scope is also stochastic: it can fall if the firm's estimate of θ rises in absolute value. Decreasing σ^2_μ causes the trend of both to be increasing through time, but the interrelation of hierarchical difficulty of tasks and focus considerations explains a stochastic process for both that is consistent with some of the key features of the data.

Since learning about θ has two effects, one as a result of the monotonic decrease in variance and the other as a result of changes in the mean, I use simulations to demonstrate the implications of the model. Ten firms are simulated for 25 periods, and the results are depicted in Figure 4. Parameters are the same as in Figure 3. The top panel shows the paths of employment and expected average costs for the 25 periods. Firms tend to grow due to increased productivity, and with the imposition of a fixed cost (here set to one), average costs tend to fall. The bottom panel describes the same simulated data, comparing growth rates of employment against the size (employment) of the firm.

As in Evans (1987) and Hall (1987), larger firms have smaller and less variable growth rates. As firms age, they learn about the θ they have drawn. The lower variance has the effect of increasing the size of firms because, as before, lower variance leads to more accurate q_s and higher efficiency. There are no predictable changes in μ; some firms learn that they have θ that is large in absolute value, and they shrink; others discover they have a small θ and grow. Recall equation (9): the weight put on the shock ϵ is $\sigma^2_\mu/(\sigma^2_\mu + \sigma^2_\epsilon)$. This falls as σ^2_μ falls over the firm's life, creating less variability as the firm ages. Because firms tend to grow as σ^2_μ falls with age, the young, variable

FIGURE 4

TIME PATHS FOR COSTS AND EMPLOYMENT
Increasing difficulty, $\phi = .15$



firms tend to be small. This is what leads to the model being able to replicate the correlation of size and growth.

Using the simulated data of the bottom panel of Figure 4, it is possible to replicate the calculation that Hall performed to estimate a model for the variance of employment growth. First, I repeat Hall's linear estimation of growth as a function of size (specifically the logarithm of employment). Then the residuals from that regression are squared and regressed on size and size squared. As in Hall, the estimation reveals a negative coefficient on size and a positive, but much smaller, coefficient on the squared term. In my model, this occurs not because very large firms have higher variability in their growth rates, but because the original model from which the residuals are generated is misspecified.

A few very small, very stable firms arise, which tends to makes the picture in Figure 4 differ from the results that Hall reports, since the model generates both very large and very small firms with stable growth rates of employment. These small, stable firms are ones that eventually learn they have drawn a large θ. The single-agent problem neglects the inherently equilibrium problem of exit that is central in the industry models that motivate the analysis here. In the Appendix I show that the informational theory of the firm developed so far can be embedded into a competitive industry equilibrium model along the lines of Hopenhayn (1992), and that existence is guaranteed. The Hopenhayn framework involves a continuum of firms so that each takes its effect on prices as negligible. Although it rules out strategic motives and the feedback they might provide to the firm's choice of scope, scale, and the learning problem, it supplies a baseline model for analyzing the interplay between firm dynamics and the organization of production within an industry. Specifically, it allows for equilibrium exit by firms with low productivity.

Equilibrium exit can reinforce the results on the model's predictions for firm dynamics. With a fixed cost F of operation, firms that discover they have a poor draw of θ exit, rather than developing into stable, small firms. Such firms, that is, firms with large $|\theta|$, must exit as their variance converges to zero whenever p is bounded and w is bounded away from zero.

Proposition 3. Suppose $\eta_s = 1 - (s(\theta + \epsilon) - q_s)^2$, $\phi > 0$, there is a fixed cost $F > 0$ any period the firm is in business, $\{p_t\}$ is bounded, and $\{w_t\} > c > 0$ for all t. Then there exists a θ^* such that any firm with $|\theta| > |\theta^*|$ chooses to exit in the long run.

Proof. In the long run, the firm's beliefs converge almost surely to $|\overline{\mu}| = \theta$, $\sigma_\mu^2 = 0$. As θ goes to infinity, scope b must go to zero, since the cost of implementing any profile $\{q_s\}$ that stays near the target $s\theta$ is prohibitively large. Since w is bounded away from zero and p is bounded above, this implies that output goes to zero as $|\theta|$ gets large. A firm facing $F > 0$ and near-zero output forever will exit to avoid paying F. *Q.E.D.*

This sort of exit reinforces the result of Figure 4, since it removes small, stable firms that arise as the long-run outcome of firms with high θ.

5. Summary

■ Evidence from panel data shows that firms vary widely both between one another and across time. A variety of industry equilibrium models have exploited this by assuming that firm productivity follows some persistent stochastic process. The learning mechanism presented here, specifically a firm learning about an unknown parameter in its technology, can account for a variety of the facts associated with firm dynamics, and can be embedded in the sort of industry model that has been used to explain a variety of industry phenomena by using stochastic, persistent firm dynamics.

The model employs the concepts of the difficulty of tasks undertaken by the firm and the ''focus'' of a firm. The model can be used to derive implications for the nature of the firm. Here, the scope of the firm is limited by increasing difficulty of tasks and the firm's limited ability to focus on a wide variety of disparate production techniques. The informational structure yields an explanation for the limits to scope. The dynamics of the learning process provide the means to compare the implications of the model to studies of firm dynamics. The data to which it is well suited can be divided into three broad categories.

First, the model explains some facts about the scope of firms. It explains why so much diversification is into related fields, why larger firms tend to be more diversified, and why firms that are more diversified tend to be larger not only overall but also at each activity they are involved with.

Second, the model can account for the learning curve, in the sense that average costs are nonincreasing over the lifetime of a firm, as many studies have found.

Third, the model can account for a variety of the stochastic properties that are suggested by empirical studies of panel datasets of firms and their implications for firm size, growth, and growth rate variability. Because the learning curve begins with fast growth and ends in a plateau, growth is negatively related to size for small firms, an empirical fact described by Sutton (1997). Despite the fact that all stochastic shocks are independent, the learning mechanism generates persistence in productivity and employment, as evidenced in Evans (1987) and Hall (1987) and used in models of turnover such as Hopenhayn (1992). Older firms have less variable growth rates and tend to be large. As a result, employment variability is negatively correlated with firm size, consistent with characteristics reported in Evans (1987) and Hall (1987). The model also makes predictions that have not been the focus of empirical research but may prove to be fruitful avenues for future work. In particular, a variety of the firm variables are conditionally heteroskedastic, and the persistence of a firm's outcome is increasing in the age of the firm.

The model has implications for other applications that model learning. For instance, Foster and Rosenzweig (1995) use a Bayesian-learning model to study data on agricultural production in India. They look at the time following the ''green revolution,'' which brought new agricultural methods to the farmers. They study the adoption decision of farmers who are uncertain about how to operate the new technology. The model presented here allows one to explore the farmer's choice of the portfolio of crops he grows, that is, the scope of agricultural production. The model predicts that scope will move with scale as farmers learn. Farmers will start out operating a few of the new-technology crops, then add to both the scale of those crops and the scope of crops they operate as they learn more about how to operate the new technology.

Several directions for further research are suggested by the shortcomings of the model in this article. First of all, the model considers only the learning dynamics of a firm learning about a single technology. Since this is a single-parameter problem with complete learning, the model predicts that eventually learning slows to the point where the firm knows θ perfectly and hence grows no further. Of course, firms are constantly adopting and adapting to new technologies. Extending the model to include multiple possible production technologies would allow, for instance, modelling the choice of firms to adopt new ideas. One could assess the role of information in the allocation of new ideas across firms. Specifically, one could compare what sort of new ideas are operated by existing firms, which ones might have some transfer of knowledge from the process they have been operating, and when it is more likely that new, specialized firms will undertake the new idea. Such a model is considered in Mitchell (1999).

Another area in which the model could be extended is in its equilibrium implications. Only the simplest equilibrium model is considered here. One possible addition would be to model each task as producing a good that entered separately into a consumer's preferences. If only a few firms could efficiently produce the most difficult items, output from that task would tend to have a high price. This sort of equilibrium consideration would add to the picture of firm dynamics introduced here. In particular, it would extend the model of firm scope to demand-side effects, since the tastes of consumers across output from the various tasks would be an important determinant of the tasks an individual firm undertakes.

Finally, the model takes firms to be single-minded entities, when in fact firms, and the knowledge produced along the learning curve, encompass a variety of people. These people must match, with each contributing his or her own knowledge to the firm. Some knowledge may be specific to the firm, and other knowledge may be applicable anywhere. Using the informational approach to learning to build this sort of model of the firm is another avenue for future research that would shed light on some important issues in firm behavior.

Appendix

■ **Existence of industry equilibrium.** I describe a competitive industry model where the firms have the learning technology described in the body of the article. Equilibrium is defined and existence is proven.

Suppose there are a large number (a continuum) of firms operating the technology each period, with either increasing difficulty, focus costs, or both. Entering firms draw a θ from a normal distribution with mean $\overline{\theta}$ and variance σ_0^2. Denote the measure defining this distribution by $N_0(\theta)$. The measure defining the shock ϵ will be denoted by $N(\epsilon)$. In order to define entry and exit, the model needs to have a fixed cost F paid by each firm for operating each period. This cost forces firms to decide between exiting and producing zero output. In addition, let there be a fixed cost E of entry. This avoids degeneracies where firms with bad draws instantly exit to be replaced by the free new draw of an entrant. Denote the measure induced by a belief $\mu = (\overline{\mu}, \sigma_\mu^2) \in \mathbb{R}^2$ by $N_\mu(\theta)$, i.e., it is generated by the open intervals of the form $(-\infty, a)$:

$$N_\mu((-\infty, a)) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{(\theta - \overline{\mu})^2}{2\sigma_\mu^2}\right) d\theta.$$

The firm's static problem of maximizing profits, net of "focus" costs, given a belief $\mu = (\overline{\mu}, \sigma_\mu^2) \in \mathbb{R}^2$, a price p, and a wage w is

$$\pi(\mu, p, w) = \max_{x_s, q_s, b} p \int_{\mathbb{R}} \int_{\mathbb{R}} \int_0^b f_s(x_s)^\gamma \left(\eta_s(q_s, \theta + \epsilon) - \phi\frac{(q_s - \overline{q})^2}{\lambda(P)}\right) ds\, dN(\epsilon)\, dN_\mu(\theta) - \int_0^b w \cdot x_s\, ds - F.$$

Denote the solution to this problem by $\{x_s^*, q_s^*, b^*\}_{s\in[0,b]}$. From the body of the article, this is a well-defined problem, it is clearly continuous, and it is bounded for any $w > 0$. Define the expected output,

$$y(\mu, p, w) = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_0^{b^*} f_s(x_s^*)^\gamma \left(\eta_s(q_s^*, \theta + \epsilon) - \phi\frac{(q_s^* - \overline{q})^2}{\lambda(P)}\right) ds\, dN(\epsilon)\, dN_\mu(\theta).$$

Let $x(\mu, p, w)$ be the associated labor demand for the firm:

$$x(\mu, p, w) = \int_0^b x_s^*(\mu, p, w)\, ds.$$

Concavity ($\gamma < 1$) and continuity imply that x_s^* is unique for each s, as well as continuous in μ, p, and w. As a result, y and x are continuous functions.

There is a single dynamically relevant action $a \in \{0, 1\}$ for the firm each period, where 1 denotes that the firm continues and 0 denotes that it exits. An equilibrium is a Borel measure $\tau_t(\mu, a)$, on $\mathbb{R}^2 \times \{0, 1\}$, that is, on the states and actions of the incumbent firms, and a mass $M_t \in R$ of firms *entering* in period t. As in Jovanovic and Rosenthal (1988), the topology of weak convergence is used for the space of Borel measures on $\mathbb{R}^2 \times \{0, 1\}$ that contains τ_t. Industry output is defined by[16]

$$Y(\tau_t, M_t) = \int_{\mathbb{R}^2\times\{0,1\}} ay(\mu, p_t, w_t)\, d\tau_t(\mu, a) + M_t y(\mu_0, p, w),$$

since continuing firms ($a = 1$) contribute $y(\mu, p, w)$, entrants contribute $y(\mu_0, p, w)$, and exiting firms contribute nothing. Likewise, industry factor demand is

$$X(\tau, M) = \int_{\mathbb{R}^2\times\{0,1\}} ax(\mu, p_t, w_t)\, d\tau_t(\mu, a) + M_t x(\mu_0, p, w).$$

There is a continuous, bounded, downward-sloping demand function $D(Y)$ for the industry's homogeneous output such that industry revenue $Y \cdot D(Y)$ is bounded, as well as a continuous, bounded, upward-sloping continuous supply function $W(X)$ with $W(0) > 0$. Define the Borel measure on the next period's mean and variance μ', which is described by equations (2) and (9) as $h(\mu' \mid \mu, \theta + \epsilon)$. For a firm in state μ, the evolution of beliefs is given by the distribution G defining the conditional Borel measure on $\mathbb{R}^2$:

[16] Here I exploit the fact that Bayes' rule is unbiased. Expected output for an individual agent with beliefs μ and action a is the same as the actual expected output for a cohort of such agents.

$$G(\cdot|\mu, a) = \int_{\mathbb{R}}\int_{\mathbb{R}} ah(\cdot|\mu, \theta + \epsilon)\, dN(\epsilon)\, dN_{\mu}(\theta).$$

For the transition G, continuity will be in the sense of the cumulative distribution function associated with the measure G, which might be denoted, as in Jovanovic and Rosenthal, $G(\mu', \mu, a)$. Across a large number of firms,[17] the distribution of states μ of active firms evolves according to its expectation, by the law of large numbers. The law of motion is given by the operator Ψ:

$$\Psi(\cdot, \tau_t, M_t) = \int_{\mathbb{R}^2\times\{0,1\}} G(\cdot|\mu)\, d\tau_t(\mu, a) + M_t G(\cdot|\mu_0).$$

The first term reflects the evolution of states for continuing firms, the latter that for entering firms. Note that this is continuous in τ_t and M_t. Finally, the dynamic program of an incumbent firm is given by

$$V_t(\mu, \tau) = \max_{a\in\{0,1\}} a\left(\pi(\mu, p(\tau_t), w(\tau_t)) + \frac{1}{1+r}\int_{\mathbb{R}^2} V_{t+1}(\mu', \tau)\, dG(\mu'|\mu, a)\right).$$

The variable a reflects exit: a firm may choose at the beginning of any period to exit ($a = 0$) and receive zero from then on. The value zero reflects the idea that an exiter who chooses to reenter does so without any of the experience gained in earlier incarnations, i.e., knowledge is lost at exit. The option value of being outside the industry is zero, since the free-entry condition ensures that entrant profits are nonpositive.

Equilibrium requires that the marginal entrant have nonpositive expected profits, that almost every firm optimize on the exit decision (which, in this case, simply amounts to the fact that it exits if its beginning-of-period value falls below zero), that supply equal demand, and that the distribution be generated by the law of motion Ψ. The initial condition is a distribution on states $\tau_0(\mu)$, where it is assumed that all firms have identical variance and a mean that is normally distributed.

Definition A1. A sequence $\{\tau_t, M_t, p_t, w_t\}_{t=0}^{\infty}$ is an *industry equilibrium* if

(i) Entry decisions are optimal:

$$V_t(\mu_0, \tau_t)\begin{cases}\leq E \\ = E, & \text{if } M_t > 0.\end{cases}$$

(ii) Exit decisions are optimal:

$$(\tau_t(\mu, a) : aV_t(\mu, \tau_t) < 0) = 0.$$

(iii) Markets clear:

$$p_t = D(Y(\tau_t, M_t)) \qquad w_t = W(X(\tau_t, M_t)).$$

(iv) Individual decisions and the aggregate distributions coincide:

$$\tau_{t+1}(\cdot, a) = \Psi(\cdot, \tau_t, M_t).$$

Theorem A1. There exists an *industry equilibrium* when either $\eta_s = 1 - (s(\theta + \epsilon) - q_s)^2$ and $\phi \geq 0$ (hierarchical difficulty of tasks) or $\eta_s = 1 - (s + \theta + \epsilon - q_s)^2$ and $\phi > 0$ (focus).

The game has been formulated as an anonymous sequential game: all the incumbents, as well as a pool of potential entrants, are players. Players produce only if they have paid F in every period since paying the entry cost E. One ‘‘state’’ is being in business or not, but that is a binary set and thus compactness and any continuity restrictions are trivial, so I ignore it. To verify existence using Jovanovic and Rosenthal’s (1988) proof of existence of equilibrium in anonymous sequential games, we need to confirm five things. The first three assumptions are straightforward to verify.

[17] Here, in addition to the fact that Bayes’ rule is an accurate representation across a large number of draws, a continuum law of large numbers is used, as in Jovanovic and Rosenthal (1987).

(i) The action space is compact. Since it is {0, 1}, it is.

(ii) The updating rule G is continuous. This is immediate from continuity of the normal distribution, continuity in the updating rules for Bayes' rule on normal distributions, and the fact that a lies in a discrete set.

(iii) The single-period reward $\pi(\mu, p(\tau_t), w(\tau_t))$ is continuous in μ and τ_t and bounded: this is true since p is bounded above and $w > 0$ because W is increasing and $W(0) > 0$, and both are continuous in τ_t.

Two assumptions are less immediate:

(iv) The total mass of players can be bounded (Jovanovic and Rosenthal normalize the mass to one, for instance).

Claim A1. The mass of players can be bounded above.

Proof. Because of the assumption that industry revenue is bounded, the industry cannot profitably support an infinite number of firms. Industrywide revenue is bounded by $\overline{R}$ in each period. If a mass of firms, call its measure S_t, are in operation, total industry profits from t on, denoted Π_t, are bounded above by

$$\Pi_t < \overline{R}\left(\frac{1}{1-\beta}\right) - (S_t \cdot F),$$

i.e., the maximum revenue $\overline{R}$ forever minus fixed costs for the S firms in operation. Since $F > 0$, there exists some $\overline{S}$ such that if $S > \overline{S}$, then total industry profits are negative from t on. Since Bayes' rule is an unbiased estimator, $\int V(\mu, \tau)\, d\tau_t$ is the total profits that will be made by all firms in operation at t over the course of their lives. But since entry is free, future entrants contribute nothing to industry profits, and so $\int V(\mu, \tau)\, d\tau_t = \Pi_t$. For all S_t bigger than $\overline{S}$, though, Π_t is negative, which implies that $V(\mu, \tau)$ must be negative for some firm operating at time t. Therefore it is impossible for S_t to ever exceed $\overline{S}$, since $V(\mu, \tau) < 0$ contradicts that all firms in operation can make at least zero from exiting. *Q.E.D.*

Finally,

(v) The state space is compact.

When we have either increasing difficulty or focus, but not both, $\overline{\mu}$ can be eliminated as a state variable. The variance of μ is contained in $[0, \mu_0]$. As a result, the state space of beliefs μ for the agents is compact, and the proof is complete.

When both $\phi > 0$ and there is increasing difficulty, $\overline{\mu}$ is a state variable, and $\overline{\mu}$ is not bounded. Construct a pseudo-industry by bounding $\overline{\mu}$ by constants $\pm\kappa_{\overline{\mu}}$ and eliminating from the industry all firms with beliefs that end a period outside those bounds. Take a sequence $\{\kappa^i_{\overline{\mu}}\}$ that approaches infinity as i grows large. Define the truncated equilibrium state-action distribution for incumbents by τ^i_t and the mass of entrants by M^i_t. Consider the sequence of prices $\{p_t, w_t\}^i$ associated with the sequence of equilibrium for the truncated economies. Since both the prices are bounded by assumption on D and S, the price sequence is an element of a compact set (in the product topology), as is M_t, by the previous argument (Claim A1). Therefore, there exists a convergent subsequence $\{p_t, w_t, M_t\}^n$. Denote the state-action distributions on this subsequence τ^n. The remainder of the proof argues that the limit of these truncated equilibria constitutes an equilibrium, since the normal distribution puts almost no weight in the tails for large n. First, it is shown that τ^n does indeed converge to some limit distribution τ.

Claim A2. The limit of the truncated economies τ^n exists.

Proof. The proof has two steps: the actions converge because the prices converge, and the states converge because the actions and the M_t's converge. Throughout, take $A = A_s \times A_a$ to be an arbitrary finite intersection of open intervals in $\mathbb{R}^2 \times \{0, 1\}$. The set of all such sets is a convergence determining class (see Stokey, Lucas, and Prescott, 1989), and so convergence on an arbitrary set of that form is sufficient to show convergence. Starting with the given initial condition $\tau_0(\mu)$, iteratively construct the action distribution from the distribution of states τ^n_s by

$$\tau^n_t(A_s, A_a) = \tau^n_s(\mu\colon \mu \in A_s \quad \text{and} \quad a(\mu, p^n_t) \in A_a)$$

and construct the state distribution $\tau^n_{t+1}(A_s, \cdot)$ according to

$$\tau^n_{t+1}(A_s, A_a) = \Psi(A_s, \tau^n_t, M^n_t).$$

Given a state distribution at time t and a sequence of prices p^n, the distribution of actions $\tau^n_t(A_s, A_a)$ converges since the decision rule a is continuous. Continuity of Ψ guarantees that the state distribution $\tau^n_{t+1}(A_s, \tau^n_t, M^n_t)$ converges as τ^n_t and M^n_t do. *Q.E.D.*

The final step shows that the limiting distribution τ is an equilibrium.

Claim A3. The limit of the truncated economies τ^n is an industry equilibrium.

Proof. For any Borel-subset E of $\mathbb{R}^2 \times \{0, 1\}$, define $\tau_t(E) = \lim_{n\to\infty} \tau_t^n(E)$ and let $M_t = \lim_{n\to\infty} M_t^n$. Continuity of Ψ directly implies that part (iv) of the definition of equilibrium is satisfied:

$$\tau_{t+1}(\cdot, a) = \Psi\left(\cdot, \lim_{n\to\infty} \tau_t^n(E), \lim_{n\to\infty} M_t^n\right) = \Psi(\cdot, \tau_t, M_t).$$

Market clearing can be shown by asserting that the limit economy has $p_t = \lim p_t^n$ and $w_t = \lim w_t^n$ and by verifying that this is an equilibrium. In that case, continuity of x and y and the fact that τ_t^n converges weakly implies

$$\lim_{n\to\infty} X(\tau_t^n, M_t^n) = \lim_{n\to\infty} \int_{\mathbb{R}^2\times\{0,1\}} ax(\mu, p_t^n, w_t^n)\, d\tau_t^n(\mu, a) + M_t x(\mu_0, p_t^n, w_t^n)$$

$$= \int_{\mathbb{R}^2\times\{0,1\}} ax(\mu, p_t, w_t)\, d\tau_t(\mu, a) + M_t x(\mu_0, p_t, w_t) = X(\tau_t, M_t)$$

and

$$\lim_{n\to\infty} Y(\tau_t^n, M_t^n) = \lim_{n\to\infty} \int_{\mathbb{R}^2\times\{0,1\}} ay(\mu, p_t^n, w_t^n)\, d\tau_t^n(\mu, a) + M_t y(\mu_0, p_t^n, w_t^n)$$

$$= \int_{\mathbb{R}^2\times\{0,1\}} ay(\mu, p_t, w_t)\, d\tau_t(\mu, a) + M_t y(\mu_0, p_t, w_t) = Y(\tau_t, M_t),$$

so $Y(\tau_t^n, M_t^n) \to Y(\tau_t, M_t)$ and $X(\tau_t^n, M_t^n) \to X(\tau_t, M_t)$. Consequently, $p_t^n \to D(Y(\tau_t, M_t))$ and $w_t^n \to W(X(\tau_t, M_t))$, and hence markets clear for τ_t, M_t, $p_t = \lim p_t^n$, and $w_t = \lim w_t^n$.

Given $p_t = \lim p_t^n$ and $w_t = \lim w_t^n$, continuity of the reward in the prices implies that $V_t(\mu, \tau_t) = \lim V_t^n(\mu, \tau_t^n)$. Continuity of V implies that parts (i) and (ii) are satisfied: if $M_t > 0$, then $M_t^n > 0$ for all sufficiently high n, and therefore $V_t^n(\mu_0, \tau_t^n) = 0$ for all sufficiently high n, so $V_t(\mu_0, \tau_t) = 0$; if $M_t \leq 0$, then $V_t^n(\mu_0, \tau_t^n) \leq 0$ for all sufficiently high n, so $V_t(\mu_0, \tau_t) \leq 0$. Finally, if $aV_t(\mu, \tau_t) = \lim aV_t^n(\mu, \tau_t^n)$ is smaller than zero, then it must also be negative for all n sufficiently large, but $(\tau_t^n(\mu, a){:}aV_t^n(\mu, \tau_t) < 0) = 0$, so therefore $(\tau_t(\mu, a){:}\ aV_t(\mu, \tau_t) < 0) = 0$.

The limit τ satisfies all of the conditions of the definition of an equilibrium. *Q.E.D.*

References

ARGOTE, L. AND EPPLE, D. "Learning Curves in Manufacturing." *Science,* Vol. 247 (1990), pp. 920–924.

AUERSWALD, P., KAUFFMAN, S., LOBO, J., AND SHELL, K. "The Production Recipe Approach to Modeling Technological Innovations: An Application to Learning by Doing." CAE Working Paper no. 98-10, Cornell University, 1998.

CARDINAL, L. AND OPLER, T. "Corporate Diversification and Innovative Efficiency: An Empirical Study." *Journal of Accounting and Economics,* Vol. 19 (1995), pp. 365–381.

ERICSON, R. AND PAKES, A. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies,* Vol. 62 (1995), pp. 53–82.

EVANS, D. "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy,* Vol. 95 (1987), pp. 657–674.

FOSTER, A. AND ROSENZWEIG, M. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy,* Vol. 103 (1995), pp. 1176–1209.

GORT, M. *Diversification and Integration in American Industry.* Princeton: Princeton University Press, 1962.

HALL, B. "The Relationship Between Firm Size and Firm Growth in the U.S. Manufacturing Sector." *Journal of Industrial Economics* Vol. 35 (1987), pp. 583–606.

HOPENHAYN, H. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica,* Vol. 60 (1992), pp. 1127–1150.

IRWIN, D. AND KLENOW, P. "Learning-by-Doing Spillovers in the Semiconductor Industry." *Journal of Political Economy,* Vol. 102 (1994), pp. 1200–1227.

JOVANOVIC, B. "Selection and the Evolution of Industry." *Econometrica,* Vol. 50 (1982), pp. 649–670.

——— AND MACDONALD, G. ''The Life Cycle of a Competitive Industry.'' *Journal of Political Economy,* Vol. 102 (1994), pp. 322–347.

——— AND NYARKO, Y. ''A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves.'' *Brookings Papers on Economic Activity, Microeconomics* (1995), pp. 247–299.

——— AND ———. ''Learning by Doing and the Choice of Technology.'' *Econometrica,* Vol. 64 (1996), pp. 1299–1310.

——— AND ———. ''Stepping Stone Mobility.'' *Carnegie-Rochester Conference Series on Public Policy,* Vol. 46 (1997), pp. 289–325.

——— AND ROSENTHAL, R. ''Anonymous Sequential Games.'' *Journal of Mathematical Economics,* Vol. 17 (1988), pp. 77–87.

LUCAS, R. ''On the Size Distribution of Business Firms.'' *Bell Journal of Economics,* Vol. 9 (1978), pp. 508–523.

MITCHELL, M. ''Adoption and the Size Distribution of Establishments.'' Mimeo, University of Minnesota, 1999.

PRESCOTT, E. ''The Multi-Period Control Problem Under Uncertainty.'' *Econometrica,* Vol. 40 (1972), pp. 1043–1058.

——— AND VISSCHER, M. ''Organization Capital.'' *Journal of Political Economy,* Vol. 88 (1980), pp. 446–461.

ROTEMBERG, J. AND SALONER, G. ''Benefits of Narrow Business Strategies.'' *American Economic Review,* Vol. 84 (1994), pp. 1330–1349.

SALTER, M.S. AND RAVENSCRAFT, D. *Diversification Through Acquisition.* New York: Free Press, 1979.

SHAW, K. ''Discussion of Jovanovic and Nyarko (1995).'' *Brookings Papers on Economic Activity, Microeconomics* (1995), pp. 300–302.

STOKEY, N., LUCAS, R., AND PRESCOTT, E. *Recursive Methods in Economic Dynamics.* Cambridge, Mass.: Harvard University Press, 1989.

SUTTON, J. ''Gibrat's Legacy.'' *Journal of Economic Literature*, Vol. 35 (1997), pp. 40–59.

WILSON, D. ''Informational Economies of Scale.'' *Bell Journal of Economics,* Vol. 6 (1975), pp. 184–195.