

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: NBER Macroeconomics Annual 1993, Volume 8

Volume Author/Editor: Olivier Blanchard and Stanley Fischer, editors

Volume Publisher: MIT Press

Volume ISBN: 0-252-02364-4

Volume URL: <http://www.nber.org/books/blan93-1>

Conference Date: March 12-13, 1993

Publication Date: January 1993

Chapter Title: How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth

Chapter Author: Ricardo J. Caballero, Adam B. Jaffe

Chapter URL: <http://www.nber.org/chapters/c10998>

Chapter pages in book: (p. 15 - 86)

Ricardo J. Caballero and Adam B. Jaffe

MIT AND NBER, AND HARVARD UNIVERSITY AND NBER

*How High Are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth**

1. Introduction and Summary

There has been a rapid growth in recent years in the theoretical literature on industrial research as an engine of economic growth.¹ At a gross level, two key concepts are at the heart of the growth process in this literature. First, profit-seeking firms try to achieve market power by producing a better good than their competitors. Over time, new goods displace old ones, earn profits for some period of time, and are then displaced in turn. This process of “creative destruction” generates the

*This paper was prepared for the NBER Macroeconomic Annual Meeting, March 14, 1993. We thank Philippe Aghion, Roland Benabou, Andrew Bernard, Olivier Blanchard, Zvi Griliches, Charles Jones, Paul Joskow, Boyan Jovanovic, Sam Kortum, Michael Kremer, Ariel Pakes, Michael Piore, Manuel Trajtenberg, and seminar participants at MIT, NYU, Georgetown, the Productivity group at the NBER, and the NBER Macroeconomic Annual 1993 meeting for their comments. Caballero thanks the National Science and Sloan Foundations for financial support; Jaffe thanks the National Science Foundation. Sam Kortum graciously shared his data on U.S. priority patents. We are particularly indebted to Manuel Trajtenberg, who created the patent citation data extract used in Section 3.2, and did so under extreme time pressure. We thank Olivier Blanchard for instigating our collaboration.

1. See Grossman and Helpman (1991a) and the references therein, in particular, Romer (1990), Grossman and Helpman (1991b), Aghion and Howitt (1992), Segerstrom (1991).

incentive for and limits the private value of industrial innovation:

. . . The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprises creates. . . . [examples] . . . [these examples] illustrate the same process of industrial mutation that incessantly revolutionizes the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism. . . . (Joseph Schumpeter, 1942)

Thus, Schumpeter recognized that innovation was the engine of growth, and that innovation is endogenously generated by competing profit-seeking firms. The second key feature of models of this process is that public-good aspects of knowledge create economywide increasing returns. In the process of creating new goods, inventors rely and build on the insights embodied in previous ideas; they achieve their success at least partly by "standing upon the shoulders of giants."² The public stock of knowledge that accumulates from the spillovers of previous inventions is thus a fundamental input in the technology to generate new ideas. This is clearly reflected in Schmookler's (1966) description of the inventor's problem:

. . . the joint determinants of inventions are (a) the wants which inventions satisfy, and (b) the intellectual ingredients of which they are made. The inventor's problem arises in a world of work and play, rest and action, frustration and satisfaction, health and sickness, and so on. . . . [I]n order to analyze the problem, to imagine possible solutions to it, to estimate their relative cost and difficulty, and to reduce one or more to practice, the inventor must use the science and technology bequeathed by the past . . .

The rich theoretical development of the growth literature can thus be seen as combining the insights of Schumpeter and Schmookler and embedding them in a general equilibrium framework. This modeling advance has not, however, been accompanied by the development of a parallel empirical literature.³ While there has been significant empirical work on different aspects of the microeconomics of technological change, there has been relatively little attempt to integrate individual micro empirical results into an overall framework for understanding growth. Our aim in this paper is to create a framework for incorporating

2. "If I have seen further (than you and Descartes) it is by standing upon the shoulders of Giants." Sir Isaac Newton, letter to Robert Hook, February 5, 1675. Newton's aphorism was popularized by Robert K. Merton, *On the Shoulders of Giants*, New York (1965).

3. A notable exception is Kortum (1993).

the microeconomics of creative destruction and knowledge spillovers into a model of growth, and to do so in such a way that we can begin to measure them and untangle the forces that determine their intensity and impact on growth.

We develop a model in the spirit of Grossman and Helpman (1992) and Aghion and Howitt (1992) that gives a simple relationship for the effect of new products on the value of existing ones. At any given time, the economy consists of a continuum of monopolistically competitive goods indexed by their quality, $q \in (-\infty, N_t]$. The newest goods are always the best, i.e., the process of research advances the frontier by increasing N_t . Because of the quality ranking implicit in this process, constant marginal cost producers see their profits—relative to those of the (new) leader—decline over time. The rate of decline depends (positively) on the degree of substitutability between new and old goods and on the pace at which new goods are introduced. This captures the endogenous process of creative destruction described earlier and, after a few algebraic steps, yields intuitive equations relating the rate of growth in a firm's value relative to that of the industry to the firm's number of new ideas relative to the industry average. By relating the concept of new ideas to that of new patents, it is possible to use these equations to gauge the empirical magnitude of creative destruction.

In order to estimate these equations, we use market value and patents data on 567 large U.S. firms. The data are annual for the period 1965–1981, and the firms are assigned to 21 technological sectors. We estimate 21 sectoral panels and find that, on average (over time and sectors), creative destruction is about 4% per year. That is, in an average sector at an average year a firm that does not invent sees its value relative to that of the industry erode by about 4%. This number varies widely across sectors; drugs has the largest average creative destruction, with about 25% per year.⁴ Because of both the endogenous variation in creative destruction and changes in estimated parameters, we also find a sharp decline in average (across sectors) creative destruction over our sample period, from a high of 7% per year in the mid-1960s to a low of 2% in the late 1970s.

Turning from Schumpeter to Schmookler, we focus on the technology by which new ideas are produced, using as inputs private research

4. We argue that, at least in part, this dispersion is due to the difficulties in measuring ideas, because patents play different roles in protecting innovation in different sectors. In other industries other mechanisms of appropriations, such as secrecy, learning curve advantages, and marketing and product support efforts are more important than patents as means of securing rents (Levin et al., 1987).

effort and the public stock of existing ideas. We focus particularly on this ideas-stock, the process by which it accumulates, and the way in which it conditions the production of new ideas.

It is well known that the standard form of the kind of “quality ladder” model that we are using embodies a strong form of research spillovers, because the same amount of resources are consumed producing the blueprint for product $q = N_t$ at time t as were consumed producing the blueprint for product $q = N_{t-dt}$ at time $t - dt$, even though the former is strictly superior to the latter. To pursue Newton’s metaphor, today’s inventors stand on the shoulders of giants that keep getting taller and never get old and weak. In order to move to a spillover formulation that can be implemented empirically, we specify how the height of the shoulders is endogenously determined by the path of previous invention.

We postulate a simple linear technology at the firm level, mapping research inputs into new ideas. This mapping changes over time as a function of the stock of public knowledge. That is, the productivity of private inputs in research varies as a function of aggregate knowledge, which is outside the control of any individual firm. We proceed to specify in some detail the process by which previous knowledge accumulates and feeds into the generation of new ideas. We postulate that it takes time for additional knowledge to diffuse sufficiently to be of use to other inventors; this tends to limit the usefulness of very recent knowledge in generating new knowledge. On the other hand, old knowledge eventually is made obsolete by the emergence of newer, superior knowledge. We call this phenomenon “knowledge” or “technological” obsolescence, and distinguish it from the obsolescence in value represented by creative destruction. That is, new ideas have two distinct effects on the current stock of ideas. They make the products represented by those ideas less valuable (creative destruction or value obsolescence), and they make the knowledge represented by those ideas less relevant in the production of new knowledge (knowledge or technological obsolescence). The strength of knowledge spillovers, and hence the growth of the economy, will depend on the parameters of the processes of knowledge diffusion and knowledge obsolescence.

At any point in time, we define the stock of knowledge available to the production of new ideas as the sum of the contribution of all previous ideas. These contributions are the product of the number of ideas in each cohort and the usefulness of the average idea in that cohort to current inventions. We describe the usefulness of an idea generated at time s for the production of new knowledge at time t ($t \geq s$) in terms of a *citation function*. In order to capture knowledge obsolescence, this

function declines with the distance between t and s in ideas-space—i.e., with the number of inventions that occur between the recipient and source cohorts. In order to capture gradual knowledge diffusion, the usefulness of old ideas increases with the calendar time between these two cohorts. We also allow for a source-cohort specific multiplicative constant that indexes the potency of the spillovers emanating from the average idea in the given cohort.

In order to estimate the citation function, we use a 1 in 100 random sample of all patents granted in the United States from 1975 to 1992, and track all their citations to previous patents back to 1900. We assume that patents are proportional to ideas and that citations are proportional to ideas used, and we estimate time-varying proportionality factors for each along with the model parameters. Our sample contains 12,592 patents and over 80,000 citations.

Several interesting findings emerge from estimating the citation function and from constructing the stock of public knowledge implied by this function. First, we find that ideas diffuse quite rapidly, with a mean lag between one and two years, which is consistent with previous estimates by Mansfield (1985) derived from survey results. Second, as postulated, knowledge obsolescence is clearly an endogenous function of the number of new ideas, rather than an exogenous function of time. The sum of squared residuals falls by about 30% when straight time depreciation is replaced by endogenous obsolescence linked to the number of new ideas. Third, the average annual rate of knowledge or technological obsolescence over the century is about 7%, but both its secular and high-frequency (endogenous) changes are quite large. It rises from about 3% at the beginning of the century to about 10–12% in 1990, with a noticeable plateau during the 1970s. Fourth, the average size of patents (measured in terms of the average number of new ideas embodied in each of these) increased over the century until the 1960s or 1970s and has declined since then. A patent in 1990 seems to contain about three times more ideas than a patent in 1900, but about 10% less than a patent in 1970. Fifth, the potency of the spillovers emanating from each cohort seems to have declined dramatically over the century: Controlling for obsolescence, we estimate that the average idea at the beginning of the century generated about five times the level of spillovers as the average recent idea. Finally, as a result of this decline in spillover potency, we estimate that the effective (or marginal) public knowledge stock declined by about 30% from 1960 to 1990, suggesting that private research productivity should have fallen by that amount.

This last result is subject to a number of caveats relating to assumptions about the exact nature of the relationship between spillovers and

citations. Its implications are, however, remarkably consonant with the data on the observed productivity of inputs in research. The observed decline in the productivity of private research, as measured by patent production, has been a subject of much research.⁵ The ratio of patents to research inputs has declined steadily since the 1950s, almost regardless of the way research-input is measured (e.g., R&D expenditures, number of scientists and engineers engaged in research).⁶ It is certainly interesting, if not surprising, that our independent measure of research productivity, which is based on knowledge flows as measured by citations, has about the same trend as directly measured productivity. Put differently, the fit of the aggregate innovation function—i.e., the function that relates aggregate (private) research inputs to total innovations—improves markedly once we include our measure of the public stock of knowledge on the right-hand side.

In the last step of the empirical section, we relate aggregate consumption growth to the rate of new idea creation. In effect, this amounts to finding the normalization constant that allows us to estimate the overall average size of a patent—a parameter that is not identified from the citation estimation alone. With this, we have empirical estimates of all of the important model parameters. Combining these estimates with a free-entry condition in the research sector and a labor market equilibrium condition, we close the model and calibrate it to fit the average rate of growth of the United States during the postwar period. The model can then be used to perform several positive and normative experiments. Although we are uncomfortable making too much of results that depend on a long sequence of assumptions and approximations, we note that the model's behavior (1) is quite consistent with the aggregate productivity slowdown in the 1970s, (2) is also consistent with the stock market boom of the 1980s (because the estimated decline in the productivity of research implies an increase in the value of existing ideas), and (3) suggests that the optimal subsidy to private R&D expenditures is around 30%.

We do not view these specific results (which are perhaps better categorized as provocative conjectures) as the main contribution of the paper. Rather, we have shown that it is possible to construct an overall modeling framework into which the key microeconomic pieces of the processes of industrial innovation and growth can be fit, and empirical estimates of the model parameters do allow the model to mimic the economy's gross growth behavior.

5. See Griliches (1989 and 1990), Kortum (1993), and Evenson (1991).

6. See Kortum (1993). Schmookler (1966) suggests that patents per research effort has been declining throughout the century.

The next pages describe the details behind this summary. We begin in Section 2 with the complete presentation of the model itself. Section 3 presents the empirical methodology and results; for reasons explained therein it is organized in a different order than this summary and the model presentation, beginning with the citation function and ending with the creative destruction equation. Section 4 calibrates the model and studies the impact of different policy and structural changes on growth and research incentives. Section 5 concludes the paper with a discussion of the overall significance of the results and suggestions for future work.

2. The Model

2.1 GOODS MARKETS

There is a representative agent endowed with a stock of ideas, \bar{L} units of labor, which have no direct utility value; an instantaneous utility function that is logarithmic in an aggregate consumption index, C_t ; and a discount factor, ρ . Using aggregate consumption as numeraire and letting r_t represent the real interest rate, we obtain the standard condition on the growth rate of consumption, \hat{C}_t :

$$\hat{C}_t = r_t - \rho. \tag{1}$$

At any point in time, the aggregate consumption index is a composite of the quality weighted output of a continuum of monopolistic competitors, which produce goods indexed by their quality: $x_t(q)$ for $q \in (-\infty, N_t]$. Quality rises monotonically over time, so newer goods are better:⁷

$$C_t \equiv \left[\int_{-\infty}^{N_t} \{x_t(q)e^q\}^\alpha dq \right]^{1/\alpha} \quad 0 \leq \alpha \leq 1. \tag{2}$$

Given aggregate consumption and the prices of each of the components of it, $p_t(q)$, consumers choose $x_t(q)$ so as to minimize the cost of that level of aggregate consumption:

$$\int_{-\infty}^{N_t} p_t(q) x_t(q) dq.$$

7. It is important to realize that the quality ladder aspect is in *addition* to the monopolistically competitive structure of the market. Stokey's (1992) elegant and general representation of preferences includes a discrete state space version of ours.

The first-order condition for this problem yields the system of demands for goods of different qualities:

$$x_t(q) = p_t(q)^{-1/(1-\alpha)} e^{(\alpha/(1-\alpha))q} C_t \quad (3)$$

At each point in time, producers take these demand functions, as well as factor (labor) prices, w_t , as given. For simplicity, let the production technology be linear and assume that process innovations have no distributive impact:⁸

$$x_t(q) = \eta_t L_t^p(q), \quad (4)$$

where $L_t^p(q)$ is labor allocated to production of $x_t(q)$ and η_t is labor productivity in the final goods sector at time t . More generally, this may be taken as the reduced form of a constant returns to scale technology including other factors of production. In the latter case the rental price of other factors would combine to add a multiplicative constant to the reduced form production function.

The linearity of technology, together with the common level of productivity and elasticities of demand faced by the infinitesimal producers of the different qualities, determines that at any given point in time all prices are identical and obey a standard markup rule:⁹

$$p_t(q) = \frac{1}{\alpha} \frac{w_t}{\eta_t}.$$

Replacing this expression in Equation (3) and the results of it in Equation (2), determines the consumption wage:

$$w_t = \alpha \eta_t \left(\frac{1 - \alpha}{\alpha} \right)^{(1-\alpha/\alpha)} e^{N_t}.$$

Thus, the price can also be expressed in terms of labor's productivity in the goods sector, η_t , and the quality level of the leading good, N_t :

$$p_t(q) = \left(\frac{1 - \alpha}{\alpha} \right)^{(1-\alpha/\alpha)} e^{N_t}.$$

8. That is, these innovations affect the technologies of goods of all qualities similarly.

9. Because of their lower quality, older goods will have smaller market shares, but because of the assumed desire for variety they never disappear completely (except in the limiting case $\alpha = 1$).

Profits accruing to a producer of a good of quality q can now be easily determined from the equilibrium values of $x_t(q)$, $p_t(q)$, and w_t :¹⁰

$$\pi_t(q) = \alpha C_t e^{(-\alpha/1-\alpha)(N_t-q)}.$$

It is interesting to notice that profits do not fall with increases in α for all levels of q . This is due to a *scope* effect. As goods become more substitutable, the profits generated by having a new—the best—good increase in spite of the reduced markup because the new product has a larger potential market. The other side of this is that goods become obsolete much faster (for a given rate of entry) because many newer goods can substitute them away: Simply put, a stronger creative destruction environment—indexed by α —is better for those that are creating and worse for those that have created in the past.

2.2 VALUATION, INNOVATION, AND LABOR MARKET

The fundamental value of a new market created at time t is:

$$v_t = \int_t^\infty \pi_\tau(N_t) e^{-\int_t^\tau \rho ds} d\tau.$$

Dividing both sides by aggregate consumption, letting $V_t \equiv v_t/C_t$, differentiating this ratio with respect to time, and recalling Equation (1), yields a differential equation characterizing the dynamic behavior of the value of an innovation in terms of units of consumption:

$$\dot{V}_t = \left(\rho + \frac{\alpha}{1-\alpha} \dot{N}_t \right) V_t - \frac{\pi_t(N_t)}{C_t}.$$

Replacing the expression for profits in this equation yields:

$$\dot{V}_t = \left(\rho + \frac{\alpha}{1-\alpha} \dot{N}_t \right) V_t - \alpha, \tag{5}$$

which is to be compared with the change in the value (in terms of units

10. Note that if the number of varieties is “small,” as is the case in the standard variety model without quality ranking where $q \in [0, N_t]$, profits would be $\pi_t(q) = \alpha C_t e^{(-\alpha/1-\alpha)(N_t-q)/(1 - e^{(-\alpha/1-\alpha)/N_t})}$. The ranking aspect of quality introduces a “discounting-like” component to the aggregators so we can work immediately with an “infinite-variety” model. This eliminates a host of short-run dynamics issues that are standard in variety models. Also see Stokey (1992).

of consumption) of the idea that has just been left behind the frontier, V_t^o :

$$\dot{V}_t^o = \rho V_t^o - \alpha. \quad (6)$$

Comparing Equations (5) and (6) shows that the “obsolescence” rate faced by owners of blueprints is $(\alpha/1 - \alpha)\dot{N}_t$, which we call the rate of creative destruction. It is proportional to the rate of advancement of the knowledge frontier. It also depends on consumers’ demand for variety; as α approaches unity, the market share of the newest product approaches unity, so we truly have a “gale” of creative destruction. One focus of our empirical work will be to provide estimates of this term for different industries and periods. We return to this later.

The other side of the value of an innovation is the cost of generating it. As is standard in the literature, we postulate a simple linear research technology at the *firm* level. A firm that invests L_{it}^r units of effort in the time interval dt generates $\theta_i L_{it}^r dt$ new blueprints.¹¹ These blueprints are worth $\theta_i L_{it}^r V_t C_t dt$ to the inventing firm; thus, free entry guarantees that:

$$w_t \geq \theta_i V_t C_t,$$

with equality if there is positive innovation.

Aggregating over all innovators yields the demand for labor by the research sector:¹²

$$L_t^r = \frac{\dot{N}_t}{\theta_t}.$$

Similarly, we can obtain the demand for labor by goods producers, L_t^p :

$$L_t^p = \int_{-\infty}^{N_t} \frac{x_t(q)}{\eta_t} dq = \frac{\alpha C_t}{w_t}.$$

11. θ_i is assumed to be deterministic at the aggregate (sectoral) level; we will model it later as a function of past knowledge accumulation in the sector. We will assume that θ_i is independent of current and previous actions by i , so the value of any particular firm is just the goods market value of its blueprints. In other words, firms do not have private stocks of past knowledge. We discuss this issue further in Section 3.2.
12. Note that θ_i may depend on aggregate quantities, including L_t^r , although in the latter case the notation is less useful.

Full employment equilibrium in the labor market is then obtained by letting:

$$L_t^p + L_t^r = \bar{L}. \tag{7}$$

2.3 SPILLOVERS, KNOWLEDGE DIFFUSION, AND KNOWLEDGE OBSOLESCENCE

The innovation function described in words above corresponds to the demand for labor in the research sector, rearranged:

$$\dot{N}_t = \theta_t L_t^r. \tag{8}$$

This equation hides in θ_t most of what is of interest to economists. It is the average productivity of research in generating new blueprints; it may contain standard *aggregate* factors of production (e.g., capital and labor)¹³ as well as spillovers from past knowledge production. We will focus on the latter but discuss briefly the former in the empirical section.

With few exceptions, the standard endogenous growth model treats θ_t as an arbitrary given constant. Such a specification conveys a strong form of intertemporal spillover, where the quality of new goods builds one for one on the top of the quality of the previous generation of goods. Labor productivity in research—i.e., θ_t —is independent of the level or pace at which ideas emerge, and is disconnected from the spillover process itself.

In this section we explicitly model several aspects of the process of diffusion of information that should influence θ_t . In particular, we consider three types of factors. First, there is the concept of endogenous obsolescence. Very old ideas are unlikely to contain much independent information that is useful for generating new ideas. Unlike the traditional notion of “depreciation,” the obsolescence of old ideas ought to be connected to the distance between ideas in the state rather than the time dimension. That is, it is not the passage of time that makes old ideas less useful, it is the accumulation of new ideas. Second, inventors take time in seeing others’ inventions, which suggests that there are diffusion lags. Unlike obsolescence, we treat the diffusion of knowledge as a function of time rather than accumulated inventions.¹⁴ Third, the

13. With either positive or negative coefficients, thus, with increasing or decreasing aggregate returns to scale in the research technology.

14. Some state dependency of knowledge diffusion is likely, but it seems plausible that time would be the primary factor.

spillover intensity between cohorts of ideas may vary independent of the effect of obsolescence of old ideas.

We capture these factors of the transmission mechanism by means of a “citations” function, $a(t, s)$ for $t \geq s$. We assume that this function depends on the probability of seeing or knowing about an idea ($t - s$) years old, and the usefulness of old ideas in generating new ones. We take the probability of seeing an idea ($t - s$) years old to be $(1 - e^{-\gamma(t-s)})$. As $\gamma \rightarrow \infty$, diffusion becomes instantaneous; $\gamma = 0$ means that all old blueprints are unavailable, so each inventor starts from scratch. In order to capture the first and last factors mentioned earlier, we assume an index of usefulness of the form $\delta e^{-\beta(N_t - N_s)}$. The term in the exponential reflects the notion that the usefulness of old ideas in the generation of new ideas depends on how far the technology has moved since the old idea. The parameter δ could capture two distinct effects. It could represent the “potency” of the spillovers emanating from each cohort of ideas. It could also represent an “absorption” parameter, measuring the intensity of use of old ideas by new ideas. The former interpretation implies that δ might vary over s ; the latter interpretation suggests the possibility of variation over t . In principle, one could imagine interaction effects, i.e., variations over (s, t) pairs. In the empirical section, we focus on variation in δ over s , i.e., variations in the potency of the spillovers emanating from different cohorts of old knowledge. There are a combination of conceptual and practical reasons for this, as will be discussed later. For now we simply treat δ as a constant, because this simplifies the explanation of the basic elements of the process of knowledge accumulation.

The citations function is the product of the usefulness of old ideas and the probability of having seen them:¹⁵

$$a(t, s) = \delta e^{-\beta(N_t - N_s)} (1 - e^{-\gamma(t-s)}) \quad t \geq s, \quad (9)$$

with $\gamma \geq 0$, $\beta \geq 0$, and $0 \leq \delta \leq 1$.

We let θ_t be the sum over all the potentially “citable” cohorts of ideas:¹⁶

15. We have saved on notation by working with a single-sector model, but it would be straightforward (from a modeling perspective) to add multiple sectors, with differing rates of obsolescence and diffusion within and across sectors. Empirical implementation of the multisector version would not be trivial. We will comment further on this in Section 5 later.

16. It is easy to add other standard ingredients to θ_t , including, e.g., decreasing returns to current labor in research. See, e.g., Kortum (1993), Stokey (1992), Jones (1992). We also comment on this possibility in Section 3.3 below.

$$\theta_t \equiv \int_{-\infty}^{N_t} a(t, s(q)) dq = \int_{-\infty}^t a(t, s) \dot{N}_s ds,$$

which can be written as:

$$\theta_t = \frac{\delta}{\beta} - \delta e^{-(\beta N_t + \gamma t)} \int_{-\infty}^{N_t} e^{\beta q + \gamma s(q)} dq. \tag{10}$$

This specification of the productivity of research effort, θ_t , has several interesting features. First, as the speed of diffusion goes to infinity, θ_t converges to a constant:

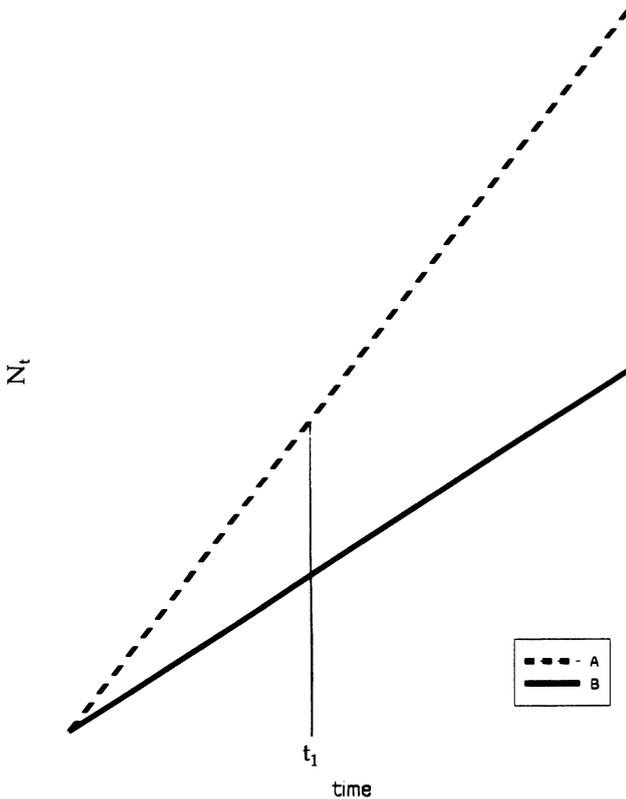
$$\lim_{\gamma \rightarrow \infty} \theta_t = \frac{\delta}{\beta}. \tag{11}$$

The insensitivity of the research productivity parameter to the rate of invention in this limiting case is the result of two offsetting factors. The increased obsolescence of the existing knowledge stock that is inherent in an economy moving (inventing) at a faster pace is exactly offset by the increased rate at which new knowledge is added to that stock. This is illustrated in Figure 1. There, we depict two economies—*A* and *B*—starting with the same level of knowledge (normalized to 0), but in *A* inventions occur at twice the rate of *B* (for reasons other than parameters of the innovation function). An inventor standing at t_1 in *A* has a larger number of inventions behind her, but the more rapid rate of invention means that a larger fraction of that stock is now obsolete. Equation (11) says that these forces exactly cancel each other when information diffusion is instantaneous, so that the marginal productivity of research in the two economies would be the same.¹⁷ Put differently, with instantaneous diffusion the right “clock” for spillovers is determined by the number of inventions: If the pace at which these occur increases, so does the speed of the economic clock, bringing about offsetting obsolescence, which leaves the amount of spillover unchanged at the margin.

Second, for given $\{\dot{N}_s\}_{s \leq t}$, θ_t is proportional to δ , which is the fraction of total knowledge that is of potential use for new inventions. Also, if diffusion is instantaneous, θ_t is inversely proportional to the rate at which new ideas outdate old ones, β . Thus, putting aside diffusion lags, the strength of spillovers depends directly on the exogenous use-

17. That is, a given amount of research labor would generate the same \dot{N} in the two economies.

Figure 1 TWO ECONOMIES WITH DIFFERENT INVENTION RATES



fulness of old knowledge, δ , and inversely on the rate at which it is made obsolete, β .

The third important feature of our formulation for θ_t is that lags in the diffusion of information—i.e., γ finite—change the relation between the pace of inventions and the productivity of labor in research by introducing a form of dynamic decreasing returns. If we return to Figure 1, if γ is finite it is no longer true that the marginal productivity of labor in research at t_1 is the same in economies A and B. Because of diffusion lags, an increase in the rate of innovation does not add to the stock of knowledge fast enough to offset the higher rate of obsolescence. The fraction of the stock of knowledge observed by inventors in an economy where the rate of inventions is relatively high is limited by the fact that a large amount of inventions have occurred only recently, when things are difficult to observe. In other words, in this case there is a second

and exogenous clock that anchors the economy.¹⁸ Thus, the productivity of research θ_t decreases with the rate of invention.

The next step in presenting the model is to describe the dynamic equilibrium behavior of the model. We postpone this until after estimating the key parameters of the model, for then the examples used to characterize equilibrium can be made more meaningful.

3. Empirical Analysis

3.1 OVERVIEW

The previous section presented a general equilibrium model of the processes of knowledge accumulation, research, innovation, product market competition, and economic growth. To estimate the parameters of the model and to test its predictions against economic experience requires finding measurable empirical constructs that correspond to the elements of the model. In this section, we plunge in and make attempts to estimate each of the important blocks of the model. We do not attempt to estimate the overall system of equations implied by the model as a whole, because the theoretical and empirical compromises that are necessary to find empirical counterparts to the model constructs cannot really be applied consistently across the different parts of the model. For example, the model has a highly stylized notion of “firms” who own no assets other than blueprints. The creative destruction Equation (5) describes the time path of the value of blueprints or ideas. To estimate this equation, we will use data on real firms.¹⁹ To do this, we will derive the model’s implications for the value of a firm, conceived as a collection of blueprints. This will involve assumptions that we believe are reasonable, but we do not go back and work out the overall implications of these assumptions for the model as a whole. Similarly, confronting the data will require us to allow for lags between invention and patent applications, patent applications and patent grants, invention and new product introduction, etc. We try to allow for these lags in

18. Although the model makes a stark distinction between lags (which occur by the “time clock”) and obsolescence (which occurs by the “invention clock”), the effects discussed here will occur as long as the speed of diffusion is less responsive to changes in the rate of innovation than is technological obsolescence.

19. The closest thing to an empirical analogue of the value of an idea is the work of Schankerman and Pakes (1986), Pakes (1986), and Pakes and Simpson (1989) on the value of patents. As these authors emphasize, however, they are estimating the value of patent protection, i.e., the *difference* between the value of the idea if it is patented and its value if it is not. Pakes (1985) estimates the stock market response to the “news” represented by a new patent. Thus, his estimates of the value of a patent exclude the portion that was predictable based on past patents and R&D.

reasonable ways, but we do not formally incorporate them into the overall model. To say it differently, we recognize that the loose correspondence between the model and the data prevents us from interpreting the model too literally.

In the following subsections we will discuss measurement issues in some detail. Overall, we will use patents as corresponding to ideas, implying the number of patents in a period, country, sector, etc. can be taken as proportional—sometime with lags—to the corresponding \dot{N} . We treat firms as agglomerations of ideas, represented by their patent holdings; we take their market value as representing the value of their idea portfolio. We use counts of Research Scientists and Engineers to represent research labor, although we explore the use of R&D expenditures as well. Finally, we use consumption expenditure from the National Income Accounts to measure total expenditure.

We present the empirical results in approximately the reverse order from the model development. We begin with the construction of θ_t , the productivity of labor in research. To do this, we use a random sample of all U.S. patents granted since 1975, and the complete history of previous patents cited by our sample patents. We take a citation as evidence that the earlier knowledge was used in the later invention, suggesting that the frequency of citation can be used to measure $a(t, s)$ in Equation (9). Because we observe many (t, s) pairs, we can estimate the parameters δ and γ of Equation (9), while at the same time estimating a (time-varying) proportionality factor between patents and $\beta\dot{N}$. From this estimation, we construct an estimate of θ_t (up to additive and multiplicative factors).

Next, we move to the innovation function, Equation (8). Using the constructed θ_t from the citation distributions, we estimate the relationship between patents and corporate research at the aggregate level in the United States. We show that by converting patents to \dot{N} using the parameter estimates from the first step, including θ_t , and normalizing the research measures in the way implied by the model, we can improve the fit between patents and research, and shed light on the puzzle noted by many researchers of the falling patent/R&D ratio in the last several decades (Griliches, 1989; Kortum, 1993). In Section 3.4, we look at the aggregate U.S. relationship between \dot{N} and the growth rate of consumption, and compare it to the prediction of Equation (21). We find that the low-frequency movements in consumption follow a pattern very similar to those in \dot{N} , although displaced in time by a few years. We conjecture that this is consistent with the model if there is a lag (not in the model) between the act of invention and the product market introduction of new goods. Finally, we return to the value side of the model. We estimate a version of Equation (5), the “creative destruction” equation,

using data on firms assigned to technological sectors. We construct estimates of the rates of endogenous obsolescence or creative destruction for these sectors during the decade of the 1970s.

3.2 KNOWLEDGE DIFFUSION, TECHNOLOGICAL OBSOLESCENCE, AND PATENT CITATIONS

As discussed in Section 2.3, the limiting form of the model has a strong form of spillovers in which the incremental innovation always comes at the same cost, regardless of how far knowledge has advanced. More realistically, inventors can build on the existing stock of knowledge, but there are limits on its usefulness in creating the next idea. Equation (10) captures the more general case in which the research productivity parameter θ_t depends on the stock of existing ideas, with each existing idea weighted by the probability that it is useful in generating new knowledge at time t . These probabilities are, in turn, dependent on the likelihood that the previous idea is known to a current inventor, and the likelihood that it is useful.

To implement this approach, we use patents as an indicator of the creation of new ideas, and the “citations” (also called references) that patents make to previous patents as an indicator of “existing ideas used in the creation of new ideas.” There is a vast literature on the virtues and vices of patent data, which addresses such issues as the large number of inventions that are never patented; variations in the “propensity to patent” of different institutions, different industries and over time; and the large variability in the “size” or importance of individual patents.²⁰ For our purposes, we will simply assume that \dot{N}_t is proportional to the rate of patenting at time t , with the proportionality factor treated as a (time-varying) parameter to be estimated.

When a patent is granted, the patent document identifies a list of references or citations, which are previous patents upon which the current patent builds.²¹ The citations serve the legal function of identifying previous patents that delimit the property right conveyed by the patent. Because citations indicate that a current invention builds on an older one, we will use the total number of citations *from* patents issued in year t to patents issued in year s as an indicator of the use of knowledge of vintage s in the production of new ideas at time t . Of course, not all citations represent spillovers; it is possible, e.g., that the inventor was

20. For a recent survey, see Griliches (1990).

21. References are also made to nonpatent materials such as scientific articles; we are not using this information. For an application that does, see Trajtenberg, Henderson, and Jaffe (1992).

not even aware of the earlier work at the time the invention was made.²² As with variations in the number of new ideas represented by the average patent, we will deal with variations in the relationship between citations and spillovers by allowing a (time-varying) proportionality factor between “ideas used” and citations, and estimating this factor as a parameter. Not surprisingly, the need to allow for this “slippage” between citations and spillovers will limit to some extent the conclusions that we can draw; we return to this issue later.

Thus, the empirical strategy of this subsection is to collect citation frequencies between patent cohorts, and use these to estimate $a(t, s)$ for many t and s . We then estimate econometrically a version of Equation (9), obtaining estimates of the parameters δ and γ , the “potency” of old ideas, and the diffusion rate of knowledge, as well as the proportionality factors that map patents into ideas and citations into “ideas used.” Producing these estimates allows us to do two things. First, we can use our estimates of the proportionality factor between patents and ideas to construct a time series for \dot{N}_t from the patent series. Second, we use the estimates of the parameters from the citation function, combined with the \dot{N}_t series, to construct θ_t , the predicted contribution of old knowledge to the production of new ideas.

Our data consist of a 1 in 100 random sample of all patents in the United States granted between the beginning of 1975 and the fall of 1992.²³ Simple statistics on these data are shown in Table 1. They consist

22. The final decision as to what citations must appear belongs with the patent examiner, but it is the result of an interactive process involving the inventor, the inventor’s attorneys, and the examiner. All of these parties can identify potential citations by searching the relevant “prior art.” Until the late 1970s, this was done by hand, using as a guide the Patent Office classification of the patent. Today, all parties have access to on-line text-search capabilities. The incentives faced by each of these parties are complicated. First, the applicant bears a legal obligation to disclose any prior art of which she has knowledge; the primary sanction for nonperformance appears to be the danger of losing the good will of the examiner (who also makes the decisions as to whether the patent will issue, what claims will be permitted, and so forth). Second, the applicant would, in a sense, prefer fewer citations, because citations may limit the scope of the property right. On the other hand, omission of important references can be grounds for invalidation of the patent, giving the applicant an incentive to make sure that citations appear. For the examiner, identifying citations not provided by the applicant is time-consuming. It appears that it is just as common for applicants and their attorneys to press for the inclusion of additional references as it is for them to resist inclusion of references (personal communication, Ms. Jane Myers, U.S. Patent Office). For more discussion on the interpretation of citations as evidence of knowledge flows, see Trajtenberg, Henderson, and Jaffe (1992), and Jaffe, Trajtenberg, and Henderson (1993).
23. Inventors from every country in the world take out patents in the United States. Of course, other countries also grant patents. We will use the phrase “patents in the United States” to refer to patents issued by the U.S. patent office, regardless of the

Table 1 PATENT STATISTICS BY CITING COHORT

Citing year	Number of sample patents	Total sample citations	Average Citations made per patent	Average lag in years	Median lag in years	Modal lag in years	Average lag in patents	Median lag in patents
1975	694	3,493	5.03	15.30	9	2	838,442	631,512
1976	689	3,352	4.87	14.67	8	2	820,938	569,471
1977	650	3,322	5.11	15.21	8	2	857,036	580,613
1978	651	3,385	5.20	14.93	8	3	846,948	578,342
1979	443	2,391	5.40	15.65	9	3	891,220	644,482
1980	648	3,690	5.69	15.67	9	4	883,639	624,897
1981	650	4,044	6.22	16.12	9	3	908,679	611,175
1982	571	3,716	6.51	16.22	10	4	927,214	677,083
1983	550	3,520	6.40	15.99	10	5	914,550	658,850
1984	662	4,058	6.13	15.39	10	3	887,513	641,609
1985	706	4,733	6.70	16.10	10	3	924,547	632,644
1986	700	4,801	6.86	16.31	11	2	952,094	704,355
1987	821	5,665	6.90	16.39	11	2	970,055	703,255
1988	766	5,487	7.16	15.77	10	3	958,933	650,782
1989	932	7,130	7.65	16.43	11	2	1,003,940	728,840
1990	928	7,458	8.04	15.35	10	3	986,169	713,108
1991	935	7,017	7.50	16.39	10	3	1,056,534	737,182
1992	596	4,515	7.58	16.46	10	3	1,082,150	761,274
All yrs.	12,592	81,777						

of 12,592 patents containing 81,777 citations. The sample varies (because of variations in the overall grant rate) from a low of 443 patents in 1979 to a high of 935 in 1991. We have valid citations going as far back as 1871.²⁴ Thus, we have observations over “*t*” from 1975 to 1992 and “*s*” from 1871 to 1992. As can be seen from Table 1, the distributions over (*t* – *s*) have extremely long tails. The mean lag in years is about 16 years; the median is about 10, and the mode is about 3.

We want to use these citation frequencies to estimate $a(t, s)$. Let $C_{t,s}$ be the observed citations in the sample from patents in year *t* to patents in year *s*.²⁵ Let S_t be the number of sample patents in year *t*, and P_s be the number of total patents in the United States in year *s*. Define

$$a^*(t, s) \equiv \frac{C_{t,s}}{S_t P_s}.$$

Thus, $a^*(t, s)$ is an estimate of the probability that a patent in year *t* cites a patent in year *s*. Panel (a) in Figure 2a shows the distribution of $a^*(t, s)$ over *s* from 1900 for each *t*. We restrict ourselves to the distributions since 1900; before that date the citation frequencies are often zero or one, and hence are very noisy estimates of the true frequency.²⁶ Panel (b) shows the function $a^*(t, s)$ for an arbitrary year (1985). The distributions shown in Figure 2a have the expected “double exponential” shape. Moreover, the increasing part is quite short, suggesting that speed of diffusion is fast. We return to this below.

nationality of the inventor or other considerations. In this subsection, we utilize a sample of all such patents. In the next subsection, we will use the phrase “U.S. patents” to mean patents (in the United States) that derive from research in the United States.

24. The citations are identified by patent number in a commercial database produced by Micropatent, Inc. Patent numbers can be used to assign grant years for the patents, because numbers are used sequentially; the patent number of the first patent issued each year back to 1836 is published in the *Historical Statistics of the U.S.* The Micropatent data contain a small but significant number (about .3%) of five-digit cited patents, which if correct would be patents issued before 1871. On inspection of the actual patent documents, we determined that many of these are, in fact, not patent numbers at all but “reissue” numbers. Thus, without manual inspection there is no way to know if any of these five-digit citations are actually valid early patents. Thus, we have simply dropped them from the dataset summarized in Table 1. Citations with six-digit or greater patent numbers appear to all be valid. Because patent number 100,000 was issued in 1870, we treat all citations 1871 or later as valid.
25. Patents are dated here by the time of grant. We will discuss timing issues further later.
26. We could, of course, estimate the variance of $a^*(t, s)$ and weight accordingly, but these estimations take very long to run as is. We decided that any additional information present in the noisy early years was not worth the increase in computational time necessary to include them.

Figure 2a $a^*[t, s]*1,000$

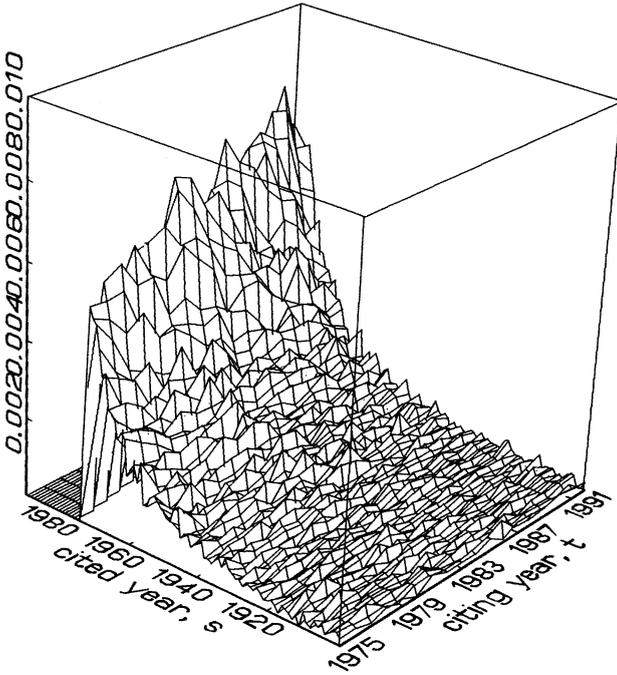
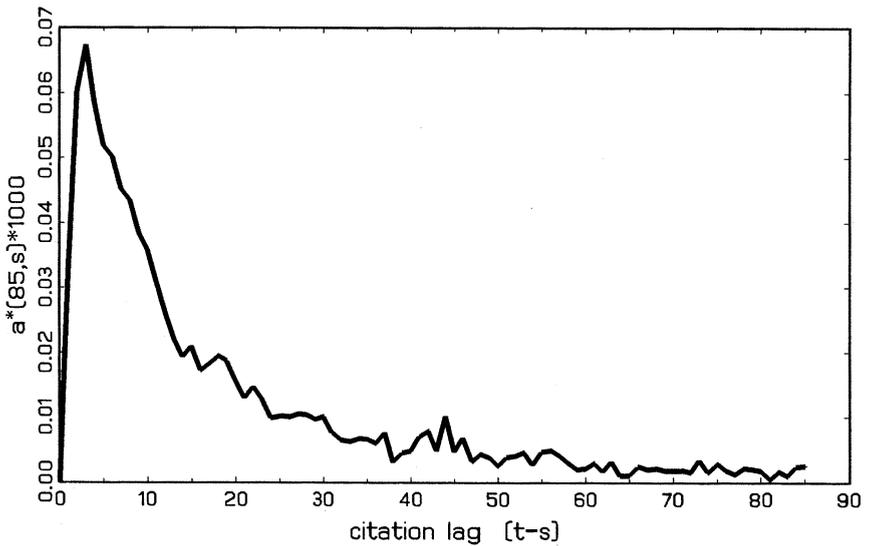


Figure 2b $a^*[85, s]*1,000$



To go from $a^*(t, s)$ to $a(t, s)$, we must be explicit about the relationships between (1) citations and “used ideas” and (2) patents and \dot{N} . We assume that citations are proportional to “used ideas” with a proportionality factor ϕ_t . That is, the patent office and its examiners have a set of rules and practices that determine what patents actually get cited. These do not affect the actual use of old knowledge in the generation of new, but they do affect the number of citations. Further, these practices can change over time. We also assume that $\beta\dot{N}$ is proportional to patents, with proportionality factor ψ_t .²⁷ We can think of ψ/β as the “average size” of a patent.²⁸ Many interpretations can be given to this “size” and its variation over time. One can think of each patent as encompassing a set of distinct ideas. Alternatively, because not all ideas are patented, one can think of ψ/β as the reciprocal of the probability that any given idea is patented. Because we care about ψ only to the extent it lets us use patents for \dot{N} , we will consider these different interpretations only to the extent that they help us think about the plausibility of the estimates. Using $C_{t,s}/\phi_t$ for “ideas used” and $(\psi/\beta)P$ for \dot{N} , we can write an expression for $a(t, s)$ in terms of observables and parameters:

$$a(t, s) = \frac{C_{t,s}}{(\phi_t/\beta^2)\psi_t S_t \psi_s P_s} = (\beta^2/\phi_t)\psi_t^{-1}\psi_s^{-1}a^*(t, s). \quad (12)$$

Because ϕ_t is purely a measurement parameter, we will absorb $1/\beta^2$ into it and simply write ϕ_t from now on. This gets us almost to the point of being able to rewrite Equation (9) (the expression for the probability that an idea will be used as a function of elapsed time and elapsed N) in terms of observables. The only additional step is to note that $(N_t - N_s)$ —the number of ideas between s and t —is, under our assumptions, just the number of patents granted between s and t , weighted by the appropriate ψ_s/β . Equation (9) can be rewritten:

$$a^*(t, s) = \phi_t \psi_t \psi_s \delta_s \exp\left(-\beta \sum_{x=s}^t \frac{\psi_x}{\beta} P_x\right) (1 - e^{-\gamma(t-s)}), \quad t \geq s. \quad (13)$$

Equation (13) is the key empirical construct of the paper. Because of the multiplicity of parameters and unfamiliarity of this sort of data, it requires several comments before we proceed to the results. First, be-

27. We choose this parameterization to emphasize that the parameter β is not identified by the patent equation. We will identify it using the growth equation below.

28. The inverse of the “propensity to patent” (Griliches, 1990).

cause of the need to estimate the proportionality factor between patents and \dot{N} , we cannot estimate β from the citation data. That is, we can use Equation (13) to recover from the citation data the *relative* size of patents in different cohorts in terms of ideas, but we cannot estimate the overall average size without bringing in additional information.²⁹ (We will use the relationship between \dot{N} and growth for this purpose.) Second, because we have multiple observations over both s and t , the parameters in this equation are all identified in principle, up to a normalization that sets one δ_s .³⁰

Third, although the parameters ϕ_t and δ_s appear symmetrically in Equation (13), we interpret them very differently. We treat ϕ_t —the proportionality factor between “ideas used” and citations—as a pure nuisance parameter, because the citations process holds no interest for us other than as a window on the spillover process. We need to allow ϕ to vary over t because citations per patent have been rising rapidly, and there are good reasons to believe that institutional changes are the reason. On the other hand, δ_s is a key model parameter; its variation over time captures changes in the potency of knowledge spillovers. As already mentioned, we find a significant fall in this potency over the century, and associate this fall with the observed reduced productivity of private research.

It is, of course, crucial for identification that we do not have parameters δ_t and ϕ_s , or δ_{st} and ϕ_{st} . That is, we do not allow the potency of spillovers to depend on the receiving cohort, we do not allow the proportionality factor between citations and “ideas used” to vary with the cited or “used” cohort, and we do not allow “interaction terms” in either. Each of these restrictions requires comment. By not allowing δ to vary over t or st , we are saying that new-invention cohorts do not vary in *their* ability to use the knowledge of the past, and that the potency of a given historic cohort in generating spillovers is a once-and-for-all attribute that does not vary over the succeeding cohorts. In other words, today’s inventors may have available to them more or less knowledge than was available to yesterday’s inventors, but there is nothing intrinsic about the nature of today’s inventions or inventive process that makes previous knowledge more or less useful to today’s inventors than yesterday’s knowledge was to yesterday’s inventors. Further, (holding obsolescence constant) the potency of, e.g., 1920 in-

29. Equivalently, we can estimate $\beta\dot{N}$ but not \dot{N} .

30. To see this, it is important to understand that ψ_s and ψ_t are not different parameters; for any given year we have the same “propensity to patent” whether we are looking at that year as a citing or cited year.

ventions for facilitating new inventions was the same in 1960 as it is today. In our model, in which quality is a unidimensional attribute so that the “nature” of inventions never really changes, these seem like natural restrictions. In a richer model, in which there were multiple quality dimensions, then one might imagine that the focus of invention today might be more or less similar to that of 1920 than the focus of invention was in 1960, suggesting that potency would vary with t and/or st . Of course, to the extent that variations in citation practices make it necessary to allow for variations in ϕ_t , it is not clear how variations in δ across t could be identified.

The restriction on ϕ , although not empty, seems more innocuous. What we are saying is that the “propensity to cite” past patents does not vary over the different historic cohorts, and that patent office practices may change over time, and this may change the number of citations (holding spillovers constant), but that these changes do not affect past cohorts differentially. Both of these propositions seem to be consistent with our impressions of the examining process. The biggest changes have been computerization of the patent data base, allowing on-line text searches to facilitate identification of citations, changes in the procedures for bringing citations to the examiners’ attention that have made it easier to include citations in the patent document, and a perceived increase in the enforcement of the legal obligation on inventors to disclose knowledge of prior art.³¹

A fourth observation of Equation (13) relates to the way the parameter ψ_t —number of ideas per patent over time—enters the equation. Because the flow of new ideas is not observed, any attempt to pin down variations in the propensity to patent requires having a second indicator (besides the rate of patenting) of the rate of knowledge generation.³² In this case, our second indicator is the rate of decline in the citation of old knowledge. That is, if the patents during some historical period were unusually large, in the sense of incorporating many ideas in each patent, then they should have made previous knowledge obsolete to a greater extent than would be expected based on the number of patents. This will be reflected in the data in the form of a reduced number of citations to these previous periods. Of course, a period with larger than average patents would also receive more citations itself, and that is captured by the presence of ψ_s in front of the exponential.³³ Because of

31. Personal communication, Jane Meyers, U.S. Patent Office.

32. See Pakes and Griliches (1984).

33. Similarly, if a period’s patents are bigger than average, they will *make* more citations; this is captured by the presence of ψ_t out front.

the presence of the δ and ϕ parameters, however, this effect probably contributes less to the estimation of the ψ s than the exponential term.³⁴

Thus, the model has two distinct parameters that relate to the average “importance,” broadly speaking, of patents of a given cohort.³⁵ The variation over time in the parameter ψ captures any differences in the number of new ideas embedded in the average patent. The variation over time in the parameter δ captures variations in the potency (in terms of spillover generation) of the ideas themselves.

Finally, we note that the diffusion of knowledge is assumed to occur at a rate that is measured in time rather than elapsed inventions. This seems natural. It is less obvious that the diffusion parameter γ need be constant over time, but we did not explore its variation.

We estimated variations of Equation (13) by nonlinear least squares on the set of observations consisting of (s, t) pairs with t varying between 1975 and 1992 and s varying between 1900 and t .³⁶ Although a model in which all of the δ s and ψ s are allowed to vary over all s and t is identified in principle, we did not attempt to estimate it. Rather, we followed a strategy of (1) always allowing a full set of multiplicative constants ϕ_i , to control for changes in citation practices, and (2) using a combination of dummies over longer time periods and polynomial functions of time to capture variations in both δ and ψ over time.

The results are presented in Table 2. The first column shows the simplest model one could imagine estimating, in which we ignore the “two clocks” and estimate both diffusion and depreciation off of the lag in years between s and t . Not surprisingly (having seen Fig. 1), this model fits the data reasonably well. We get an estimate for γ of about .8, and an estimate for the “obsolescence” rate of about .075 per year. As would be expected from the rising average citations made per patent shown in Table 1, the estimates of ϕ_i rise from 1975 to 1992. This is a result that is apparent in all specifications. Next, we substitute elapsed patents for time in the depreciation term, while still maintaining con-

34. If we estimated the model with a free and complete set of the parameters δ_s , there would be no contribution to the estimation of the time pattern of ψ from its presence out front. Because, however, we constrain the δ_s to follow particular functional forms, this is not the case.

35. Note that the “size” of ideas themselves, in terms of the product quality improvement they allow, does not vary except in the specific way defined by the exponential form in which q enters the aggregate consumption good (Equation (2)).

36. Because the $a^*(t, s)$ are estimated and the frequencies differ greatly, the model is heteroskedastic. We did not deal with this problem explicitly, but dropping the early observations can be interpreted as limiting ourselves to that part of the data in which the heteroskedasticity is likely to be less. The standard errors reported are heteroskedasticity consistent, however.

Table 2 CITATION FUNCTION REGRESSION RESULTS

Parameter	1	2	3	4
γ	.816 (.019)	.703 (.015)	0.707 (0.012)	0.705 (0.016)
ψ	.074 (.001)	.062 (.001)	a	a
ϕ_{1975}	1.000 —	1.000 —	1.000 —	1.000 —
ϕ_{1992}	1.244 (.028)	1.444 (.027)	1.495 (0.012)	2.066 (0.021)
MSE	0.184	0.130	0.124	0.122
2 (LLK-LLKpc)	—	537.6	70.6	15.0

Notes: Dependent variable: Sample citations from year t to year s / ((Sample patents) $_t$ · (total patents) $_s$).
Sample: t from 1975 to 1992; s from 1900 to t .

*See Figure 2.

Estimates of $\phi_{1976} - \phi_{1991}$ are omitted to conserve space. (LLK-LLKpc): log-likelihood minus the log-likelihood of the previous column.

stancy over time in δ_s and ψ_s . To facilitate interpretation of the results, we use for the terms in the summation in Equation (13) the number of patents in each year divided by the average (over the whole sample, 1900–1992) number of patents per year. This makes the parameter in front of the term $(N_t - N_s)$ the average annual obsolescence rate; it is therefore directly comparable to the time-obsolescence rate estimated in Column 1. Estimating obsolescence based on patents rather than time improves the fit markedly,³⁷ and also reduces the average obsolescence rate to just over 6% per year. Because the number of patents is greater in recent years, the observed prevalence of early citations is consistent with a lower average annual obsolescence rate than when the rate is held constant over time.

The third column of Table 2 “frees up” the parameter ψ to vary over both t and s , i.e., it allows for variations in the propensity to patent over time (while still keeping the spillover potency of ideas constant over time). Needless to say, there are many different ways to represent the movement in ψ_t . We explored a number of these, and they generally give similar overall results. The version reported in Column 3 of Table 2 models ψ_t with a single dummy for the years 1900–1919, a second dummy for 1920–1939, a third dummy for 1940–1959, and a cubic equation in the log of t for the period 1960–1992. This improves the fit further, and the parameter estimates are quite significant. The time path of ψ implied by these estimates can be seen in Figure 3a. Generally

37. The sum of squared residuals is reduced by about 30%.

Figure 3a: ψ_t

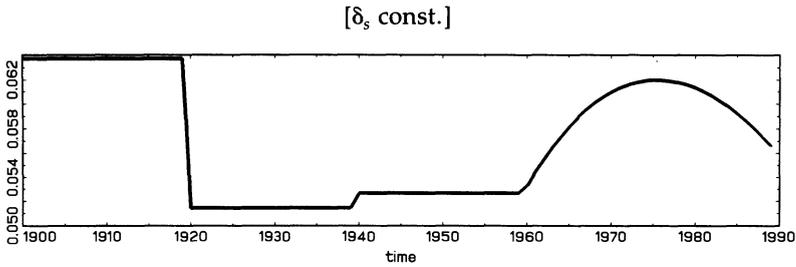
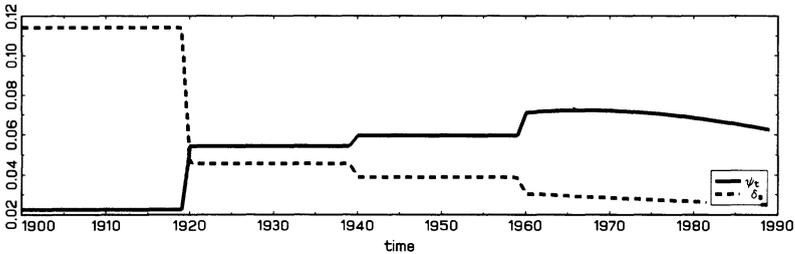


Figure 3b: ψ_t and δ_s

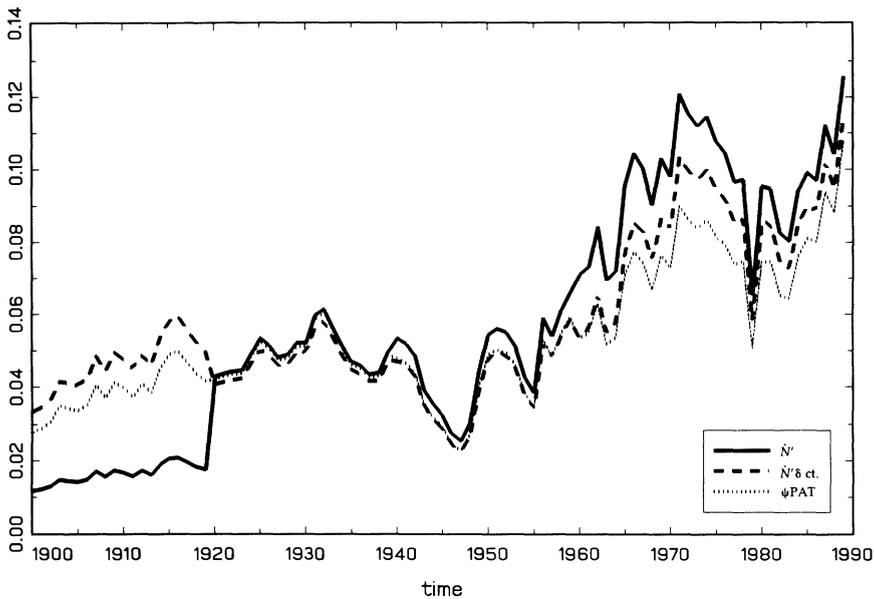


speaking, the path rises over the century, reaching a peak somewhere during the 1970s, and then begins to decline. Again, the patent counts have been divided by the average patents per year so that the magnitude of ψ can be interpreted as the annual rate of obsolescence created by an average year's worth of patents.³⁸

Column 4 builds on Column 3 by freeing up δ_s . The parameterization of δ_s is parallel to that for ψ_s , with dummies for long periods early in the century and a cubic equation in t for the period 1960–1992. This yields a similar pattern for ψ_t to what we had before, except in the very beginning of the century. But δ_s moves significantly in the opposite direction, as shown in Figure 3b, falling significantly from the start of the century until about 1960, and then leveling off into a slower decline. As we will see later, the decline in δ_s shown in Figure 3b translates into a secular decline in the predicted productivity of research, θ_t . In other words, knowledge from successive patent cohorts over the century is being incorporated in current patents at rates that imply that the potency of later cohorts in facilitating new knowledge generation is mark-

38. The number of patents per year also changes over time, of course, causing the variation in the yearly rate of obsolescence to be much greater than the variation in ψ_t . See Figure 4.

Figure 4: TECHNOLOGICAL OBSOLESCENCE



edly less than the potency of earlier cohorts. Because more recent cohorts get more weight (they are less obsolete) in current knowledge, the predicted effective spillover rate (and hence research productivity) falls over the century.³⁹

As noted earlier, the estimate of the diffusion parameter γ is not very sensitive to these specification issues. It is consistently about .7 to .8, suggesting an average lag until knowledge has diffused of between one and two years.

For obsolescence, it is not ψ_t that matters, but rather $\psi_t P_t$, which is equal to $\dot{N}' \equiv \beta \dot{N}$. Figure 4 shows different estimates of \dot{N}' , compared to the overall patent series itself. What the picture shows is that first, the variations in ψ over time are small relative to the movements in patents. Nonetheless, the “corrected” series does show a noticeably different pattern, particularly at the beginning of the century and from the end of World War II until the late 1970s. In this latter period, our estimate of \dot{N}' increases almost 40% more than the patent series itself. After the early 1970s, ψ_t begins to decline, exacerbating the fall in the

39. One manifestation of this phenomenon is the presence of fat tails in the distribution of the $a^*(t, s)$'s. This is not enough, however: allowing for fat tails in estimation improves the fit but it leaves—to a large extent—unaffected the declining path of δ_s .

rate of patenting itself that occurs between 1970 and the early 1980s. Then patenting picks up again, and although ψ_t is still falling, \dot{N}' picks up as well. In the next subsection, we turn to a more detailed analysis of trends in \dot{N}' versus trends in patents.

The last output of the citations analysis is the construction of the series θ_t , our estimate of the productivity of labor in research. From Equation (10), θ_t is the integral over all past ideas q of $a(t, s(q))$. We do not observe $a(t, s)$, but the estimated citation equation can be used to construct predicted values of $a(t, s)$, using the parameters γ , δ_s , and ψ_s and the data series P_t . This is easily done by replacing Equation (13) in (12).

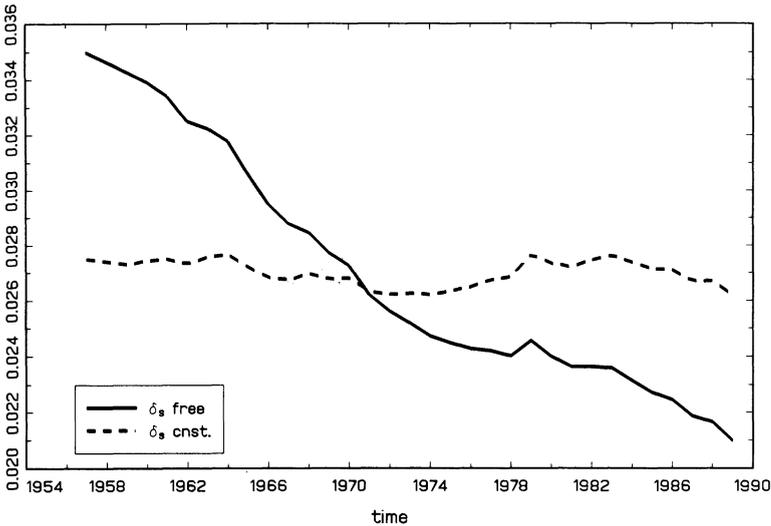
Our estimate of θ_t (up to a constant) is then easily obtained from a discrete representation of the definition of θ_t :⁴⁰

$$\theta_t \approx \sum_{s=0}^t a(t, s) \psi_s P_s.$$

In the formulation described previously, in which ϕ_t enters the relationship between $a(t, s)$ and $a^*(t, s)$ but does not affect $a(t, s)$ itself, the parameters ϕ_t do not enter into the construction of $a(t, s)$ or θ_t . We also explore a variation in which we interpret the parameters ϕ_t as representing something real about the use of knowledge rather than a citation artifact. This will change the estimated path of θ_t after 1975.

Two potential estimates of θ_t from the citation function are plotted in Figure 5. The solid line corresponds to Column 4 of Table 2, in which δ_s is allowed to vary over time. It shows a dramatic fall in the predicted productivity of research labor, very rapid from the 1950s to the early 1970s, and then somewhat slower than that. The heavy dashed line in Figure 5 corresponds to Column 2 of Table 2, i.e., it holds δ_s constant over time. It shows a much flatter pattern of research productivity. In the next subsection, we will relate the estimated θ_t to the observed productivity of research in the United States. For now, it is important to emphasize that this time series is not generated from data on the productivity of research. Rather, it is the model's *prediction* about the path of research productivity, based on the pattern of old knowledge used, as represented by citations, in the production of new knowledge. What is driving the trend is the path of δ_s . In a nutshell, the citations data show that recent cohorts of patents are less cited than older ones

40. The fact that the summation starts from 0 rather than minus infinity is empirically irrelevant because the first t we study is sufficiently large (60) so the value of the excluded $a(t, s)\dot{N}$, is negligible.

Figure 5: THE STOCK OF PUBLIC KNOWLEDGE (θ)

(controlling for obsolescence), suggesting that they are less potent in generating spillovers. Because obsolescence makes recent patents more important in the overall stock, the current stock is less potent overall than the stock that was available to previous inventors. With shorter shoulders to stand on, current inventors have to spend more on telescopes in order to see as far as their predecessors did.

Note that the estimated decline in θ_t is conditional on our assumption that the parameter ϕ_t captures only citation behavior and not any change in the actual use of old knowledge. If, on the other hand, one believed that the increase in the raw citation rate that can be seen in the data is a real (exogenous) increase in the use of old knowledge, then we would expect this increase to feed through into rising research productivity. It seems likely, a priori, that the large increase in citation intensities reflects primarily a change in citation practices. In addition, as we will show later, actual research productivity shows no evidence of increasing after 1975 as would be predicted if θ_t were rising steeply.

3.3 THE INNOVATION PRODUCTION FUNCTION

Equation (8) describes the production of innovations as a function of the research labor force L_t^r and a research productivity function or parameter θ_t . In the previous subsection we have developed a method for constructing an estimate of θ_t based on the "use," as evidenced by citations, of older knowledge. In this subsection we will incorporate this

estimate into estimates of the innovation function itself.⁴¹ We estimate the innovation function on aggregate time series for patents and two measures of research inputs—R&D spending and research scientists and engineers—for the period 1957–1989. If the data and model are interpreted literally, Equation (8) leaves large serially correlated disturbances unexplained. One possibility is to correct for serial correlation, leaving this dynamic pattern in the disturbance unexplained. Another possibility is to modify the theory so innovations are a direct function of current and lagged research. Doing the latter modifies our model only slightly if the lagged research that matters is the aggregate one, while it makes the theory more cumbersome if lagged research is private. From the point of view of estimation in this section, however, this distinction does not matter. Moreover, this common specification is indistinguishable from a third explanation where the serially correlated disturbance is attributed to the timing of research, innovation, and patenting. We explain and adopt the latter, but it should be clear that we have no strong position on the relative importance of these sources of serial correlation.

We will treat the fundamental innovation equation (8) as holding with respect to unobserved new ideas. These ideas do not, however, lead instantaneously to patent applications. Rather, patent applications P_t are given by:

$$\psi_t P_t = \dot{N}_t^{ob} = \frac{(1 - \rho)\dot{N}_t'}{1 - \rho L}.$$

Thus, as previously, we allow for a time-varying propensity to patent or proportionality constant between ideas and patents; we call this ψ_t , and we will use the estimates from the previous subsection to convert P_t to \dot{N}_t^{ob} . In addition, however, we allow for lags in the conversion of ideas into patent applications. We will estimate these lags, parameterized by ρ , from the innovation function itself. We take the actual productivity parameter, $\hat{\theta}_t$, to depend on the θ_t estimated earlier and exogenous research productivity: $\hat{\theta}_t = \eta_0 + \eta_1 \theta_t$. The parameters η_0 and η_1 will also be estimated from the innovation function.

Note that patents are not actually granted until some later date, usually within two to three years of application but occasionally much later.

41. We will also use the estimates of ψ_t from the previous subsection to convert patents to \dot{N}' . Given the large inflow of foreign patents, this is likely to underestimate the change in size of U.S. patents, for on average there will be more inventions in between subsequent U.S. patents.

Because this second lag is variable and results from the vagaries of the patent office, we estimate the innovation function using patents by year of application.⁴² This is in contrast to our construction of θ_i , and the knowledge diffusion analysis more generally, which used patents by grant year. This was predicated on the assumption that knowledge does not begin to spread until the patent is actually granted. This seems plausible, because patent applications are secret. Only when the patent is granted is the technical knowledge contained in it published. We should note, however, that we will look below at the response of firms' market value to (ultimately successful) patent applications. We are implicitly assuming that, at the time of application, the market knows that an idea has been generated, and responds to that knowledge, even though its technical content is still secret.

We estimate the innovation function using measures of U.S. research inputs, and a measure of U.S. patents. Again, this differs from the previous subsection where, though we are using "patents in the United States," we include patents granted in the United States to foreigners in \bar{N} . This means that, in estimating the relationship between U.S. research and U.S. patents, we include in the spillover function θ_i all patents, not just U.S. patents. Thus, we are assuming that U.S. research produces U.S. inventions, but it draws upon (and is made obsolete by) worldwide inventions.

It is well known that the productivity of research, as measured by patent output, shows a long-term decline from the 1950s until the mid-1980s (Griliches, 1989; Kortum, 1993). This is shown in Figure 6. The top panel shows the ratio of patents to several measures of research input; the bottom panel plots \bar{N}' , i.e., ψ_i times patents. The patent series is total "U.S. priority" patents,⁴³ by year of application. The research input measures include real nongovernment R&D expenditures and total research scientists and engineers, as well as each of these scaled by U.S. population,⁴⁴ and nominal R&D scaled by nominal expenditure. Explanations that have been put forward for the downward trend in patent productivity include (1) an exogenous fall in "technological opportunity," (2) aggregate decreasing returns to research, producing a fall in average productivity because research has risen significantly, and (3) a decline in the propensity to patent (Kortum, 1993).

42. This is the standard practice in the patent literature. See, e.g., Hausman, Hall, and Griliches (1984).

43. This means that the patent was applied for in the United States before being applied for anywhere else in the world.

44. Civilian population over the age of 16 (1991 Economic Report of the President).

Figure 6a: PATENTS/RESEARCH

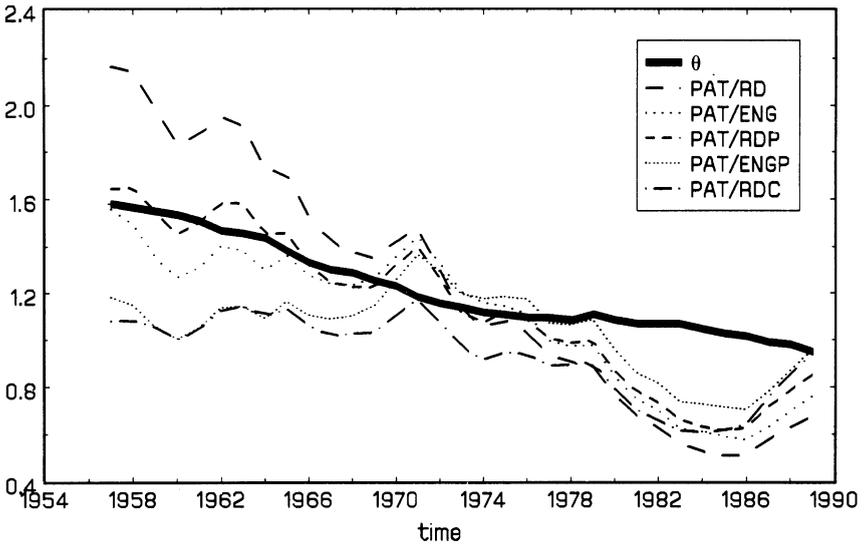
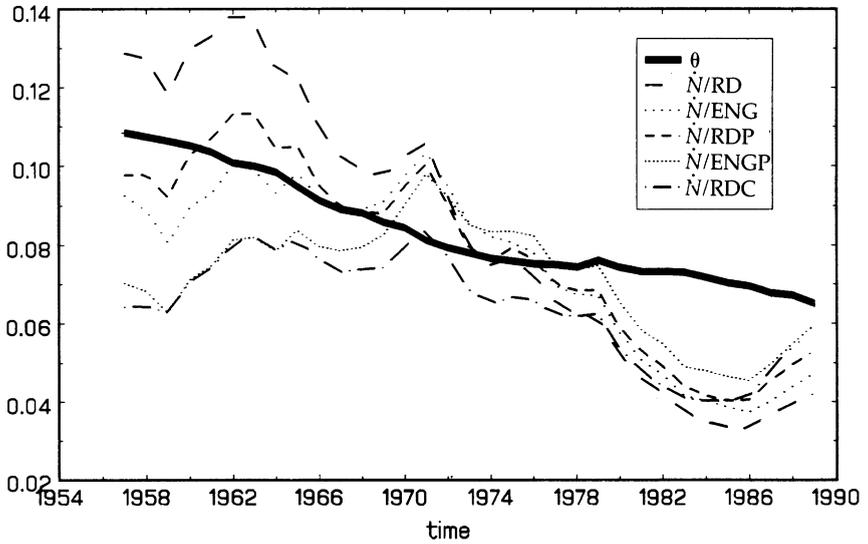


Figure 6b: \dot{N} /RESEARCH



Our estimates from the previous section shed significant light on these issues. First, as can be seen from Figure 6 (as well as Fig. 4), correcting for patent size using the estimated ψ_t does mitigate the fall in productivity up until 1970. Thereafter, unfortunately, the estimated ψ_t begins to fall, aggravating the apparent fall in productivity. Our estimates for θ_t do, however, provide an explanation for much of the overall trend in patent productivity. This can be seen from Figure 6, in which the estimated θ_t is plotted along with the observed productivity. In both panels, it is clear that the overall downward movement in θ_t is quite consistent with the fall in research productivity, although it does not explain the high-frequency movements, including the precipitous drop in the late 1970s and the rapid rise in very recent years. In the terms of the previously offered explanations for the fall in patent productivity, our estimates suggest that "technological opportunity" has indeed fallen. In our model, this takes the form of decreased usefulness of the stock of existing knowledge in generating new ideas.⁴⁵ The previous section shows that this fall can be observed in the pattern of actual use of older knowledge, as evidenced by patent citations.

Figure 6 suggests that the estimated θ_t explains much of the observed trend in patent productivity. To push this a little further, we estimate the equation:⁴⁶

$$\dot{N}_t^{ob} = \alpha_0 + (1 - \rho)\bar{\theta}_t R_t + \rho \dot{N}_{t-1}^{ob},$$

with R_t a measure of research input and

$$\bar{\theta}_t = \eta_0 + \eta_1 \theta_t.$$

The parameter estimates are presented in Table 3. The columns correspond to different measures of research input. In column 1, we use research scientists and engineers. The fit is quite good, and the esti-

45. Note that the θ_t shown in Figure 6 is the one that results when we treat the increase in ϕ_t as an artifact of citation practices rather than a real phenomenon. On the one hand, the close correspondence of the resulting θ_t to measured productivity provides further support to our conjecture that the movements in ϕ_t are not "real." On the other hand, if this is wrong and the "abnormal" trend in citations corresponds to a true increase in spillovers, our measure of θ_t exacerbates rather than eliminates the patent/R&D ratio puzzle, at least until 1986.

46. We also estimated versions allowing for decreasing returns with respect to research input. The standard specification with decreasing returns but $\eta_1 = 0$ was uniformly and very significantly outperformed by the linear model with η_1 unrestricted. Adding decreasing returns to the model with η_1 unrestricted yielded unrealistically low and very imprecise estimates of the returns to scale parameter.

Table 3 INNOVATION FUNCTION RESULTS

Parameter	S-E 1	S-E/Pop. 2	R&D 3	R&D/C 4	R&D/Pop. 5
α_0	-0.012 (0.008)	-0.013 (0.007)	0.002 (0.009)	-0.009 (0.062)	-0.013 (0.007)
ρ	0.953 (0.069)	0.934 (0.062)	0.887 (0.076)	0.912 (0.074)	0.934 (0.062)
η_0	-0.380 (0.134)	0.002 (0.056)	-0.133 (0.133)	-0.042 (0.052)	0.028 (0.056)
η_1	0.759 (0.223)	0.259 (0.071)	0.188 (0.189)	0.207 (0.104)	0.254 (0.071)
LLK	180.0	182.4	176.6	177.9	178.6

Notes: Dependent variable: ψ_t weighted aggregate U.S. priority patents by year of application. Sample: 1958–1989.

mates are all reasonable and statistically significant. As suggested earlier, θ_t is highly significant. The next column uses research scientists and engineers as a fraction of the population. The fit is approximately the same, and the role of θ_t is smaller but still positive and significant. In the next three columns, we report results for research input measured as real R&D expenditure, and R&D expenditure divided by consumption and population, respectively. Except for unscaled R&D (where the signs are correct but the coefficients are not significant), the results are similar to those obtained with scientists and engineers.

Thus, the regression results confirm what can be seen in the pictures, that our estimated decline in θ_t , inferred from patent citations, “explains” much of the secular decline in measured patent productivity. In interpreting this, we must consider the factors determining the almost monotonic decline in θ_t through our sample period. First, there is the decline in δ_s , indicating a reduction in the usefulness of successive cohorts of ideas in generating spillovers to the creation of new ideas.⁴⁷ In principle, there is a second force potentially at work: θ_t is constructed using all patents, not just U.S. patents. The fraction of U.S. patents going to foreigners rose from about 11% in 1957 to about 44 percent in 1989. From the point of view of U.S. inventors, this increase in foreign patenting in the United States has the effect of speeding up the

47. The empirical regularity is that the citations to early patents are more frequent than would be expected based on the estimated rate of exponential obsolescence. We interpret this in terms of δ_s having been larger in the early years. Alternatively, one could say that the true obsolescence function is “slower” than exponential, i.e., the citation distributions have fatter tails than predicted by exponential obsolescence. Either way, the effect is similar; we would predict a decline in the effective spillover base as knowledge accumulates.

“ \dot{N} clock” without affecting the “time clock.” New ideas are coming faster in the aggregate, making it harder for any inventor to take a step, and much of this new knowledge is too recent to have diffused and thereby spilled over to helping new invention.

The top panel of Figure 7 shows that it is actually only the decline in δ_s that mattered. The solid line shows what θ_t would have looked like if δ_s had been constant; it is itself quite constant. Figure 7 also shows why the increase in \dot{N} caused by foreign patenting did not matter: The rate of knowledge diffusion is fast enough so that the spillovers from this influx roughly balanced the increased obsolescence. This can be seen from the dashed line, which shows what θ_t would have looked like if γ were much smaller, i.e., .001. In that case, we would have had a marked decline in θ_t even if δ_s had been constant. The bottom panel reproduces these two cases for the actual (declining) path of δ_s . It shows that, if γ had been smaller, there would have been an additional downward effort on productivity from the influx of foreign patenting. But, given the actual γ , this effect is small; diffusion is close enough to instantaneous that we are, in effect, in the world described in Section 2.3 in which θ_t does not depend on \dot{N} .

3.4 \dot{N} AND GROWTH

As shown in Equation (21), the theoretical model predicts an extremely simple linear relationship between the growth rate of consumption and

Figure 7a: θ FOR DIFFERENT γ WITH δ_s CONSTANT

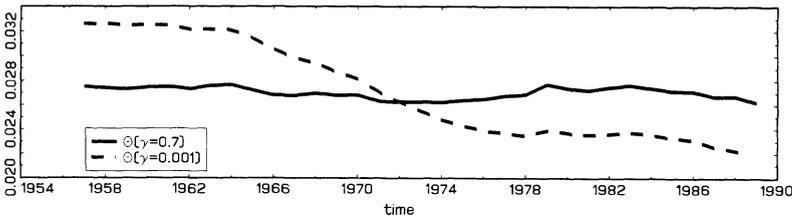


Figure 7b: θ FOR DIFFERENT γ WITH δ_s FREE

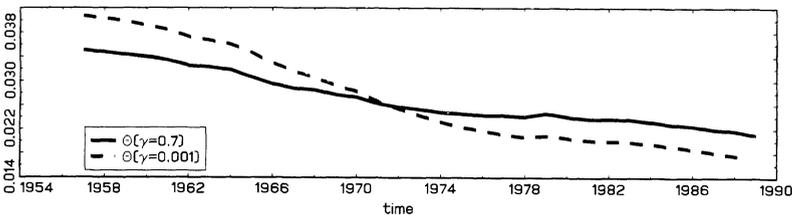


Figure 8a: GROWTH AND INNOVATION

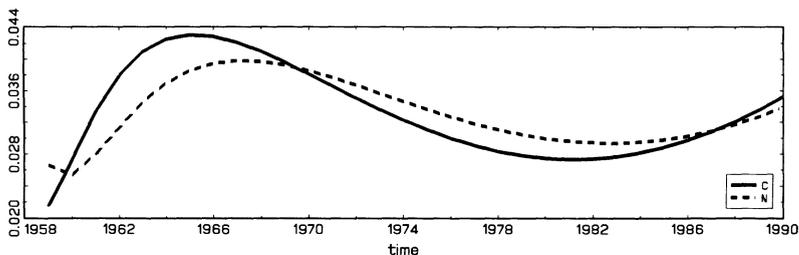
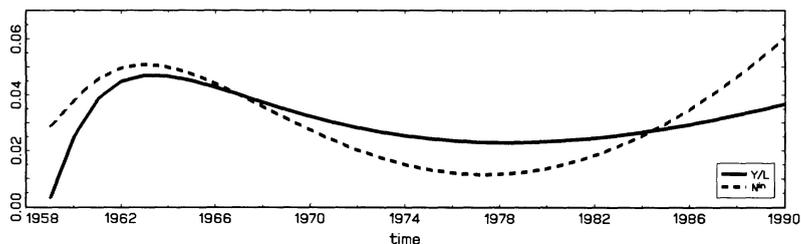


Figure 8b: GROWTH AND INNOVATION



\dot{N} . Casual inspection of the data makes clear that such a relationship does not hold for annual data in the United States. The high-frequency movements in these series are not likely to be well explained by a growth model. Therefore, to explore whether we can find evidence of the predicted relationship, we smoothed both time series by using predicted values from a regression of the actual series on a fifth-order polynomial in the log of time. The top panel of Figure 8 shows the resulting smoothed consumption growth rate and \dot{N} , using the same U.S. priority patent series, corrected by the estimated ψ_t , from the citation data. The shapes are strikingly similar, especially considering that it is not clear that one can expect consumption, as actually measured in the National Income Accounts, to move as predicted by the model.⁴⁸

Given the previous discussion, it is not clear how seriously one should take precise timing issues. For completeness, however, we mention that the \dot{N} series appears to be displaced forward by one or two

48. The essence of technological change in this model is the introduction of new goods. As has been emphasized by Griliches (1979) and others, the extent to which the statistics capture the increase in consumption that occurs when new goods are introduced varies greatly across industries. The authorities measure revenues, not output, and convert revenues to output using price deflators that generally ignore the quality improvement associated with new goods.

years up until the early 1980s. This suggests either that new ideas are incorporated in new products even before the date of patent application, or, perhaps more likely, that both series are moved by other shocks but exhibit different dynamic responses to these.

From Equations (2) and (4), it is possible to write:

$$\frac{C_t}{\eta_t} = \left[\int_{-\infty}^{N_t} \{L_t^p(q) e^{\theta q}\}^\alpha dq \right]^{1/\alpha},$$

but because

$$\int_{-\infty}^{N_t} L_t^p(q) dq = \bar{L} - \frac{\dot{N}_t}{\theta_t},$$

and

$$L_t^p(q) = L_t^p(N_t) e^{(\alpha/1-\alpha)(q-N_t)},$$

we can express the rate of growth of consumption as:⁴⁹

$$\hat{C}_t \approx \hat{\eta}_t + \frac{1}{\beta} \dot{N}_t' - \Delta \left(\frac{\dot{N}_t'}{\beta \bar{L} \theta_t} \right).$$

We estimated the following empirical version of this equation:

$$\hat{C}_t = \lambda_0 + \lambda_1 \dot{N}_t' - \lambda_2 \Delta \left(\frac{\dot{N}_t'}{\theta_t} \right). \quad (14)$$

The coefficient λ_2 was never significant, so we omit the last term in the regressions reported below in Table 4. Columns 1 and 2 present results for the growth rate of consumption, with and without a serial correlation correction. Columns 3 and 4 present the same results using the growth rate of labor productivity instead of consumption as the dependent variable. All versions tell a similar story. The coefficient on \dot{N}' is about .5 to .6 and significant.⁵⁰

The bottom panel of Figure 8 shows the (smoothed) growth rate of

49. For this we use the approximation $\Delta \ln(1 - x) \approx -x$, for x small.

50. All coefficients appear significant, but our transformation introduces large biases in the standard errors, so these should not be taken too seriously. Again, we only emphasize the coincidence in the general shape of the curves in Figure 8.

Table 4 GROWTH EQUATION REGRESSION RESULTS

Parameter	1	2	3	4
λ_0	-0.0092 (0.0044)	-0.0098 (0.0032)	-0.0248 (0.0132)	-0.0311 (0.0083)
λ_1	0.6121 (0.0629)	0.5999 (0.0434)	0.5162 (0.1872)	0.5440 (0.1128)
ρ	—	0.9037 (0.042)	—	0.8993 (0.0362)
R^2	—	—	—	—
LLK	5.752	7.237	4.697	6.315

Notes: Dependent variable: smoothed growth rate of U.S. consumption expenditure. Sample: 1958–1989.

labor productivity (GNP over employment), and the “true” \dot{N} that can be derived from \dot{N}^{ob} using the estimated parameters from the innovation equation. Again, the movements are very closely related. Although we stress that the lag we have incorporated between the true and observed \dot{N} is something of a black box, the model does seem to do a good job at predicting the longer-term movements in the productivity series.

3.5 CREATIVE DESTRUCTION

All of the previous empirical subsections can be thought of as conditional on the path of research. In the model, the allocation of labor to research is determined by the value of new ideas, whose time path is given by Equation (5), the “creative destruction” equation. In this subsection, we present some empirical estimates of that equation.

As noted earlier, estimation of this equation requires confronting the notion of firms. It also requires identifying the concept of sectors, which have not been explicitly described in the models but whose dynamic properties can be easily understood by extension of the results from the single-sector model. We will treat firms as agglomerations of blueprints, although we will not seek to explain why any particular firm holds the particular portfolio of blueprints that it does.⁵¹ We will assign firms to sectors, which will be defined as groups of firms whose research activities have historically focused on similar areas. With these assumptions, we can derive a version of the creative destruction equation that relates the deviations from the sector mean in firms’ value growth rates to the deviations from the mean of the firms’ \dot{N} .

51. Although this definition of firms is consistent with the nonexcludability of knowledge implicit in the model, it is unlikely to hold true in reality. In other words, research know-how, organizational capital, and other forms of private knowledge must add value to a firm beyond the value of its patents.

Let F_{its} , I_{ts} and I_{ts}^s represent the value of a firm i in sector s , the value of the entire sector, and the value of the firms in sector s that are included in the sample; all of them at time t and in terms of units of consumption. Letting $\Lambda_i(q)$ and $\omega_{ts}(q)$ be indicator functions, we have:

$$F_{its} = \int_{-\infty}^{N_{ts}} \Lambda_i(q) V_t(q) dq,$$

$$I_{ts}^s = \int_{-\infty}^{N_{ts}} \omega_{ts}(q) V_t(q) dq,$$

$$I_{ts} = \int_{-\infty}^{N_{ts}} V_t(q) dq.$$

Differentiating these expressions with respect to time, using Equation (5), letting $\dot{N}_{its} \equiv \Lambda_i(N_{ts}) \dot{N}_{ts}$, and assuming $\omega_{ts}(q) \approx \omega_{ts}$, we obtain our basic estimating equation:⁵²

$$\hat{F}_{its} - \hat{I}_{ts}^s = \lambda_{ts} \frac{I_{ts}^s}{F_{its}} \left[\dot{N}_{it} - \frac{F_{its}}{I_{ts}^s} \dot{N}_{ts}^s \right], \quad (15)$$

where

$$\lambda_{ts} \equiv \frac{\alpha_s}{1 - \alpha_s \omega_{ts}}.$$

We estimate Equation (15) on an unbalanced panel of firms from the NBER R&D panel (Hall et al., 1988), which contains Compustat financial information and U.S. patent data. The assignment of these firms to technological sectors is described in Jaffe (1986). Briefly, the distribution of the firms' patents across patent classes for the period 1965–1972 was used in a multinomial clustering algorithm to identify groups of firms with "similar" patent class distributions. The 567 firms are assigned to a total of 21 sectors. Simple statistics for the sectors are presented in Table 5. In general the level of aggregation of the sectors is comparable

52. An alternative derivation of the same equation can be obtained by letting $\Lambda_i(N_{ts})$ be a random variable independent across i , so the best predictor of its realization is the share of the firm's value in the industry. Also, if one assumes that each sector is comprised of a large number of firms, the total number of new patents in the industry together with its change in value can be taken as known in advance (or at least uncertainty about these can be assumed to be negligible relative to the same concepts at the firm level).

Table 5 STATISTICS FOR CREATIVE DESTRUCTION SAMPLE FIRMS

Sector	Number of firms	Total observations	Average firm value	Average patents per firm	Average growth rate of value	Firm patents times sector/ firm value	Average sector value	Average sector patents	Average estimated rate of creative destruction
1	30	441	542.23	5.7642	0.04455	352.45	14735.46	167.02	0.0145
2	44	684	2351.03	67.0877	0.0743	3768.71	96900.08	2889.46	0.0318
3	16	226	4070.11	78.0487	0.03546	4622.25	57903.42	1141.43	0.0129
4	21	318	2652.53	45.1792	0.05299	1123.76	49921.29	907.29	0.2511
5	16	239	706.93	16.841	0.05748	337.34	9956.7	255.97	0.0304
6	20	303	8224.99	79.8119	0.05127	5477.57	146432.80	1544.65	0.0018
7	24	341	208.75	5.0469	0.07579	143.5	4158.79	112.46	0.0411
8	21	326	4530.83	58.7791	0.14140	3378.6	88065.11	1209.72	-0.0012
9	33	489	1286.94	54.1984	0.09234	2063.89	38265.89	1708.56	0.1313
10	27	418	1607.58	57.2847	0.12077	2159.74	39483.43	1508.15	0.0693
11	27	393	1147.94	34.916	0.02499	1349.37	27557.81	877.26	0.0246
12	34	511	1071.48	6.3053	0.03260	226.17	32952.44	205.77	0.0115
13	31	451	502.24	14.6386	0.08774	695.64	13135.86	418.41	0.0056
14	13	200	1726.71	15.34	0.08214	475.53	20197.74	195.03	0.0065
15	33	493	1333.17	16.7728	0.04794	872.42	39768.57	525.17	0.0105
16	23	342	547.36	9.462	0.07767	518.28	11241.32	207.74	0.0205
17	49	757	2145.09	41.749	0.06716	3995.33	99554.18	2018.69	0.0205
18	24	339	587.77	9.6962	0.04060	285.85	11881.87	208.21	0.0084
19	29	425	751.29	12.4965	0.05136	931.02	19417.10	346.84	0.0101
20	27	393	300.32	3.4122	0.10359	318.63	6767.87	85.57	0.0152
21	25	368	632.95	15.9484	0.05136	653.64	13540.95	375.9	0.0297
All sectors	567	8,457	1658.93	31.3948	0.06210	1712.52	43774.37	940.05	0.0355

Notes: Sectors: 1. Adhesives and coatings; 2. Chemicals; 3. Electrochemistry; 4. Drugs; 5. Cleaning and abrading; 6. Petroleum and refining; 7. Machinery (non-elec.); 8. Computers and data processing; 9. Electrical equipment; 10. Electronic communications; 11. Stone, clay, and glass; 12. Food; 13. Instruments; 14. Medical; 15. Primary metals; 16. Misc. consumer goods; 17. Automotive; 18. Paper and packaging; 19. Refrig. and heat exch.; 20. Static structures; 21. Farm and construction equipment.

to two- to three-digit SIC industries. The assignment is made, however, on the basis of areas of inventive activity rather than sales.

To estimate Equation (15), we need to parameterize the variation in the parameter λ_{st} over s and t . This parameter encompasses variations in the CES parameter α , in the share of the sector represented by the firms in the sample, and also variations in the proportionality factor between patents and new ideas. We treat it as the product of a sector-specific constant and a cubic polynomial in t . We constrain λ to be positive by using an exponential time polynomial.⁵³ Although Equation (13) implies that the two terms in square brackets are constrained to have the same coefficient λ_{st} , we allow a free parameter on the sector patent total \dot{N}_{ts}^s . We also allow for year- and sector-specific intercepts, leading to the equation actually estimated:

$$\hat{F}_{its} - \hat{I}_{ts}^s = \alpha_{ts} + \lambda_t \lambda_s \left[\dot{N}_{it} \frac{I_{ts}^s}{F_{its}} - \mu \dot{N}_{ts}^s \right], \quad (16)$$

The results of estimating this equation on 8,457 observations are presented in Table 6. The coefficients λ_s are generally positive, although many are not significant.⁵⁴ The parameter μ , which should be unity if the proportionality (implied by the model) between value and patents holds, is about 1.4. This says that firm patents scaled by the ratio of firm to sector value averages less than sector patents. This is consistent with the general and intuitive finding that large firms have proportionally fewer patents than small firms.⁵⁵ The parameters P_1 , P_2 , and P_3 in the table are the coefficients of the cubic time polynomial for λ_t .

To interpret these results, we use the parameter estimates to calculate rates of creative destruction. The most straightforward way to do this is to multiply the estimated λ_{st} times the estimated μ times the number of patents in the sector in each year. Doing this yields estimates of the rate of creative destruction by sector by year. The average over the sample years of these numbers are presented in the last column of Table 5. They range from essentially zero for a number of sectors, including computers, to a high of 25% per year for drugs. The (unweighted) average across all sectors is about 3.5% per year. Some aspects of these results are quite consistent with previous findings. In particular, the

53. If we do not constrain these estimates to be positive, we obtain negative estimates at the end of the sample, although these are insignificant. The overall fit was statistically unaffected by our nonnegativity constraint.

54. The time and sector intercepts are not generally significant.

55. See, e.g., Bound et al. (1984).

very high rate of creative destruction for drugs is consistent with the general view that this is a very progressive sector *and* one in which patents are a very good measure of technical advance (Mansfield, 1985; Levin et al., 1987). We also find relatively fast creative destruction as measured by patents in machinery, electrical equipment, and communications equipment. These are all sectors where patents are reasonably important. In contrast, our inability to find creative destruction in computers is probably related to the relative unimportance of patents in that sector (Bound et al., 1984; Levin et al., 1987), rather than a low rate of technological change.

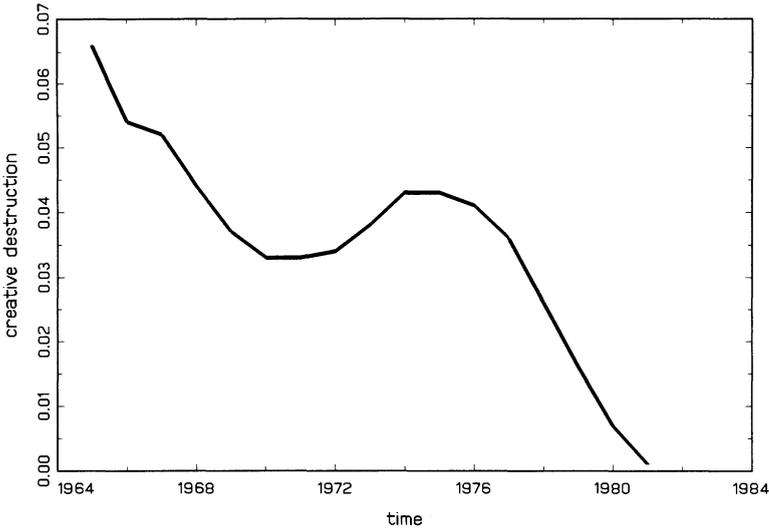
We can also look at variations over time. Again, the most straightforward way to do this is to simply multiply the estimated λ_{st} and μ times

Table 6 CREATIVE DESTRUCTION REGRESSION RESULTS

<i>Parameter</i>	<i>Coefficient</i>	<i>Standard error</i>
μ	1.383	.245
λ_1	.318	.185
λ_2	.037	.025
λ_3	.040	.019
λ_4	1.034	.297
λ_5	.425	.420
λ_6	.004	.009
λ_7	1.397	.767
λ_8	-.003	.030
λ_9	.283	.089
λ_{10}	.157	.046
λ_{11}	.098	.060
λ_{12}	.187	.238
λ_{13}	.045	.087
λ_{14}	.135	.178
λ_{15}	.072	.071
λ_{16}	.376	.177
λ_{17}	.038	.018
λ_{18}	.146	.271
λ_{19}	.108	.063
λ_{20}	.695	.249
λ_{21}	.316	.158
P_1	-.808	.206
P_2	.113	.034
P_3	-.0048	.0016
$R^2 = .0366$		
$MSE = .1612$		

Notes: Dependent variable: firm value growth rate minus sector value growth rate. Sample: 8,457 observations on 567 firms, 1966-1981.

Figure 9: CREATIVE DESTRUCTION



the yearly sectoral patent total. If we do this, and average over sectors, we get the path shown in Figure 9. Beginning at a high of about 7% in 1965, creative destruction falls quickly into the range of 3–4%, and then falls close to zero at the end of the sample period in 1981. There is, however, reason not to take the time variation in total patents in these data too seriously. First, it is affected by the changing firm composition in the unbalanced panel. In addition, total patents in this sample fell precipitously in 1980 and 1981, because of the way the data set was created.⁵⁶ For these reasons, the very low rates at the end of the sample period should probably be ignored.

4. General Equilibrium, Calibration, and Implications of the Empirical Results

In the previous section we used the basic structure of the growth model presented in Section 2 to guide our search for empirical manifestations of creative destruction and knowledge spillovers. In this section we go back to the model itself and examine its properties, using the estimates

56. Recall that the data is patents by year of application. Because the data set was created in 1982, some ultimately successful applications from 1980 and 1981 had not yet been granted, leading to a systematic undercount in those years.

obtained in the empirical section for the parameter values. The primary purpose of this section is simply to explore the static and dynamic behavior of the model using reasonable parameter values. We will also, however, go a little further and examine some strong positive and normative conjectures that arise from the behavior of the model when calibrated with the empirical parameter values.

In Section 2 we identified the following key parameters: ρ , α , δ , \bar{L} , γ , and β . Section 3 provides estimates of α , γ , and β , as well as of changes in δ (but not its level) and in α over time. Initially, we focus our attention on the average value of the parameters, and postpone the discussion of the impact of changes in parameters until later in this section. We set the discount rate, ρ , to 0.03,⁵⁷ and use average U.S. consumption growth together with the steady state of the model to calibrate δ and \bar{L} . In order to calibrate these parameters, we first need to go back to the model itself and characterize its equilibrium.

The dynamical system that emerges from the model described in Section 2 has a range of parameters for which innovation is unprofitable, so growth does not occur. We focus our analysis on cases where steady-state growth is strictly positive.

From the innovation function, labor market equilibrium and free-entry conditions, we obtain an expression for the rate of innovation as a function of the productivity of labor in research and of the value of the leading idea in units of consumption:

$$\dot{N}_t = \theta_t \bar{L} - \frac{\alpha}{V_t}. \quad (17)$$

Replacing this in the valuation equation (5) yields the dynamic equation for V_t as a function of itself and θ_t :

$$\dot{V}_t = \left(\rho + \frac{\alpha}{1 - \alpha} \theta_t \bar{L} \right) V_t - \frac{\alpha}{1 - \alpha}. \quad (18)$$

Finally, the dynamic equation for labor productivity in research is obtained by differentiating Equation (10) with respect to time:

$$\dot{\theta}_t = \gamma \left(\frac{\delta}{\beta} - \theta_t \right) - \beta \theta_t \dot{N}_t,$$

57. Quantitative conclusions are not affected by other “reasonable” assumptions about the discount rate, ρ .

which combined with Equation (17) yields:

$$\dot{\theta}_t = \gamma \left(\frac{\delta}{\beta} - \theta_t \right) - \beta \theta_t \left(\theta_t \bar{L} - \frac{\alpha}{V_t} \right). \quad (19)$$

Equations (18) and (19), together with initial conditions on θ and a transversality condition, form a self-contained dynamical system. After solving for the paths of θ_t and V_t from this system, the rate of innovation can be recovered from Equation (17).

Because we found large values of γ —i.e., a high speed of diffusion of ideas—in the previous section, it is convenient to first characterize the case where diffusion is infinitely fast; this is a good approximation, and it has the advantage of an extremely simple set of dynamic equations.

If there are no lags in the diffusion of knowledge, the system has no transitional dynamics.⁵⁸ As shown in Section 2, in this case $\theta_t = \delta/\beta$; which by Equation (18) and the transversality condition implies:

$$V = \frac{\alpha}{1 - \alpha} \times \frac{1}{\rho + \frac{\alpha}{1 - \alpha} \frac{\delta}{\beta} \bar{L}},$$

while the rate of innovation is:

$$\dot{N} = (1 - \alpha) \left(\frac{\delta}{\beta} \bar{L} - \rho \right), \quad (20)$$

and consumption growth is:

$$\hat{C}_t = \dot{N} + \hat{\eta}_t. \quad (21)$$

These expressions provide a simple setup to understand the main role of α , δ , β , and ρ in determining the equilibrium valuation of new ideas, knowledge spillovers, and the economy's rate of growth.⁵⁹

58. Obviously, anticipated changes will lead to non-steady state dynamics. The absence of transitional dynamics refers to the response of the system to a once-and-for-all unexpected change in a constant of the model.

59. One could also study the impact of \bar{L} , but we take this as a nuisance parameter. It is at best unclear which is the appropriate normalization.

$$\frac{\partial V}{\partial \alpha} = \frac{\rho V^2}{\alpha^2} > 0; \frac{\partial \dot{N}}{\partial \alpha} = -\frac{\dot{N}}{1-\alpha} < 0; \frac{\partial}{\partial \alpha} \left(\frac{\alpha \dot{N}}{1-\alpha} \right) = \frac{\dot{N}}{1-\alpha} > 0, \quad (22a)$$

$$\frac{\partial V}{\partial \delta} = -\frac{\delta}{\beta} \frac{\partial V}{\partial \beta} = -\frac{\bar{L} V^2}{\beta} < 0; \frac{\partial \dot{N}}{\partial \delta} = -\frac{\delta}{\beta} \frac{\partial \dot{N}}{\partial \beta} = \frac{(1-\alpha)\bar{L}}{\beta} > 0, \quad (22b)$$

$$\frac{\partial V}{\partial \rho} = -\frac{(1-\alpha)V^2}{\alpha} < 0; \frac{\partial \dot{N}}{\partial \rho} = -(1-\alpha) < 0. \quad (22c)$$

When the degree of substitutability among goods (α) rises, the value (per unit of consumption) of a new idea rises. This may seem surprising because an increase in α lowers the markup charged by firms. There are, however, three other effects that must be considered. First, as discussed in Section 2, the fall in the markup is outweighed by an increase in the size of the market faced by new ideas (the “scope effect”), so that the initial profit of the newest idea rises with α . Second, an increase in α raises creative destruction, which reduces expected future profits and, hence, the initial value of ideas. Third, it can be shown that from these effects alone, the ratio of the value to the wage would fall.⁶⁰ From the free-entry condition, this would be inconsistent with positive invention. Therefore, there must be an endogenous decline in creative destruction (fall in \dot{N}) in order to offset the fall in the value to wage ratio.⁶¹

The impact of an increase in the potency of spillovers (δ) as well as that of a reduction in the technological destructiveness of new ideas (β) is shown in Equation (22b). They increase the pace of innovation, and through the impact of this on creative destruction, lower the equilibrium value of new ideas.⁶² Finally, Equation (22c) shows that an increase in consumers’ impatience, ρ , lowers both the value of new ideas and the rate of invention through standard discounting and savings mechanisms.

Although the intuition as well as the sign of the relations described

60. An important mechanism behind the monotonic relation between growth and markups is that labor supply is completely inelastic. If this assumption is relaxed, then as the wage falls (i.e., markups rise), there would be a reduction in resources available and, under the appropriate functional assumptions, an eventual decline in equilibrium growth.

61. In the γ finite case, the endogenous decline in creative destruction would not completely offset the initial decline in the value to wage ratio.

62. Alternatively, the fall in equilibrium value can be explained in terms of the increase in the productivity of research. This and the creative destruction interpretations of the decline in value are related in equilibrium by the free-entry condition.

earlier survive the introduction of a finite γ ,⁶³ it is worth describing briefly the implications of frictions in the diffusion of ideas.

If γ is finite, the system exhibits transitional dynamics because “the clocks have to synchronize to the new pace.” That is, if information diffuses slowly, “shocks” that lead to changes in \dot{N} disrupt the balance between technological obsolescence and increases in the base of knowledge. Transitional dynamics occur while the new level of θ that restores this balance is reached. Before discussing dynamics, however, it is worth pausing to study the steady state and to calibrate the remaining parameters using average U.S. growth data.

The steady state can be found in closed form, although the equations are somewhat less informative than before:

$$\theta = \frac{\rho(1 - \alpha)\beta - \gamma + \sqrt{(\rho(1 - \alpha)\beta - \gamma)^2 + 4\delta\gamma\bar{L}(1 - \alpha)}}{2\beta(1 - \alpha)\bar{L}}, \quad (23a)$$

$$\dot{N} = \frac{\sqrt{(\rho(1 - \alpha)\beta - \gamma)^2 + 4\delta\gamma\bar{L}(1 - \alpha)} - \rho(1 - \alpha)\beta - \gamma}{2\beta}, \quad (23b)$$

and

$$V = \frac{\alpha}{1 - \alpha} \times \frac{1}{\rho + \frac{\alpha}{1 - \alpha}\theta\bar{L}}. \quad (23c)$$

It is apparent from these equations that using average growth data only (which we do here), it is neither possible nor relevant to separate \bar{L} from δ ; thus we set $\bar{L} = 1$. We can now recover all the parameters of the model. We obtained $\gamma = 0.7$ directly from the citation function, and $\beta = 1.67$ is the inverse of the coefficient on the change in the number of ideas, as normalized in the citation function, in the growth equation. We recover α from the average of our creative destruction estimates, 0.035, which corresponds to $\alpha\dot{N}/(1 - \alpha)$, and the average of \dot{N} , 0.042.⁶⁴ The estimate of α so obtained is 0.463. The last parameter, δ , is obtained from the steady-state equilibrium equation for \dot{N} (Equation 23b) and is equal to 0.199.

63. This is particularly true for large values of γ , as is the one estimated in the empirical section.

64. For this we use that $\dot{N} = \dot{N}^{\text{ob}}/\beta$, and $\dot{N}^{\text{ob}} = 0.07$. Our sample for the estimate of creative destruction is 1965–1981, while we use the period 1960–1989 to compute the average change in ideas.

Figure 10a plots the steady-growth rate for an economy with the same base parameters of the United States and a range of values of α , the index of creative destruction, and δ , the spillover potency index, that contain the U.S. values. Figure 10b does the same for the equilibrium value/consumption ratio. U.S. “average” equilibrium is depicted by a black dot in each figure.

One of our main empirical findings is that productivity of labor in research has declined sharply over the sample, and this seems to be mostly due to a decline in δ . According to Figure 10a, this ought to lower the equilibrium rate of innovation, \dot{N} , and raise the value of a new patent to consumption ratio, V . On the other hand, our empirical evidence on creative destruction suggests that α has decreased over time; this should raise \dot{N} and V .

Splitting the sample into two periods, 1960–1974 and 1975–1989, associating the 1965–1974 and 1975–1981 averages of creative destruction to each of these periods, respectively,⁶⁵ we can calculate the model’s predicted steady-state changes in \dot{N} and V . We find that the effect of the decline in the power of spillovers dominates the effect of the decline in creative destruction on equilibrium growth, leading to a prediction that \dot{N} should have fallen by about 50% from the first to the second periods. With respect to value, both of these effects go in the same direction, leading to a predicted increase of about 25% in the value to consumption ratio.

In reality, \dot{N} (the patent series adjusted by our estimated ψ) fell about 15%. If we proxy the value to consumption ratio by the ratio of stock prices to nominal consumption, we find an actual rise in V of about 20%.⁶⁶ Thus, the qualitative predictions of the model are confirmed, although the actual magnitudes changed less than the model implies they should have.

We conclude this section by briefly addressing several issues that are tangential to our main concerns: (1) a description of the transitional dynamics of the model, (2) the long-run effect of changes in the speed of diffusion of ideas (γ) and in technological destructiveness (β), and (3) optimal R&D subsidy rates.

Figure 11 shows the phase diagram corresponding to a case with

65. Remember that the sample used to estimate the path of creative destruction goes from 1965 to 1981 only.

66. There are several reasons to think that an index of aggregate stock prices is not a great proxy for the value of patents. In particular, the number of patents a firm has is likely to be an important component of the value of its stock and, for the experiments we discuss here, value and number of patents at the firm level are likely to be negatively correlated.

Figure 10a: \dot{N}

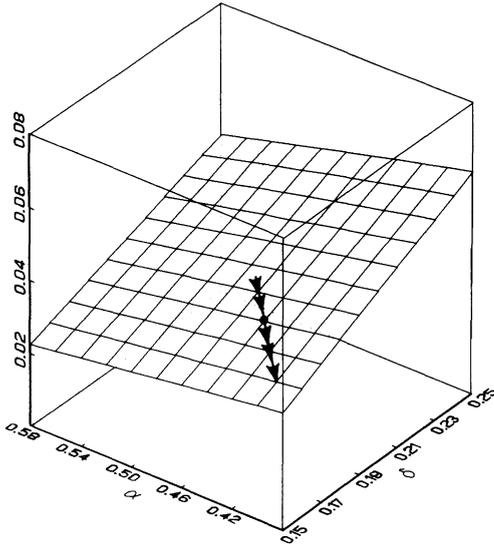


Figure 10b: V

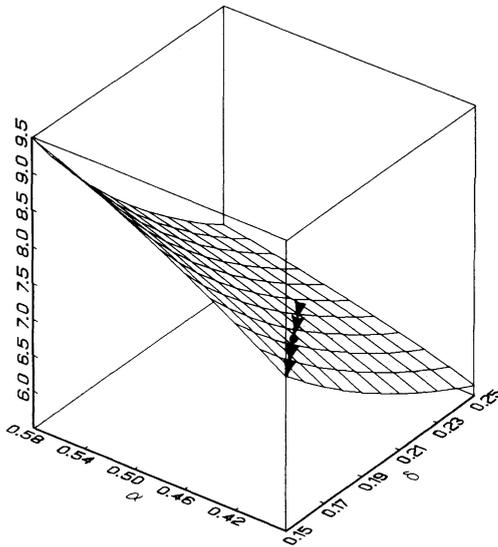
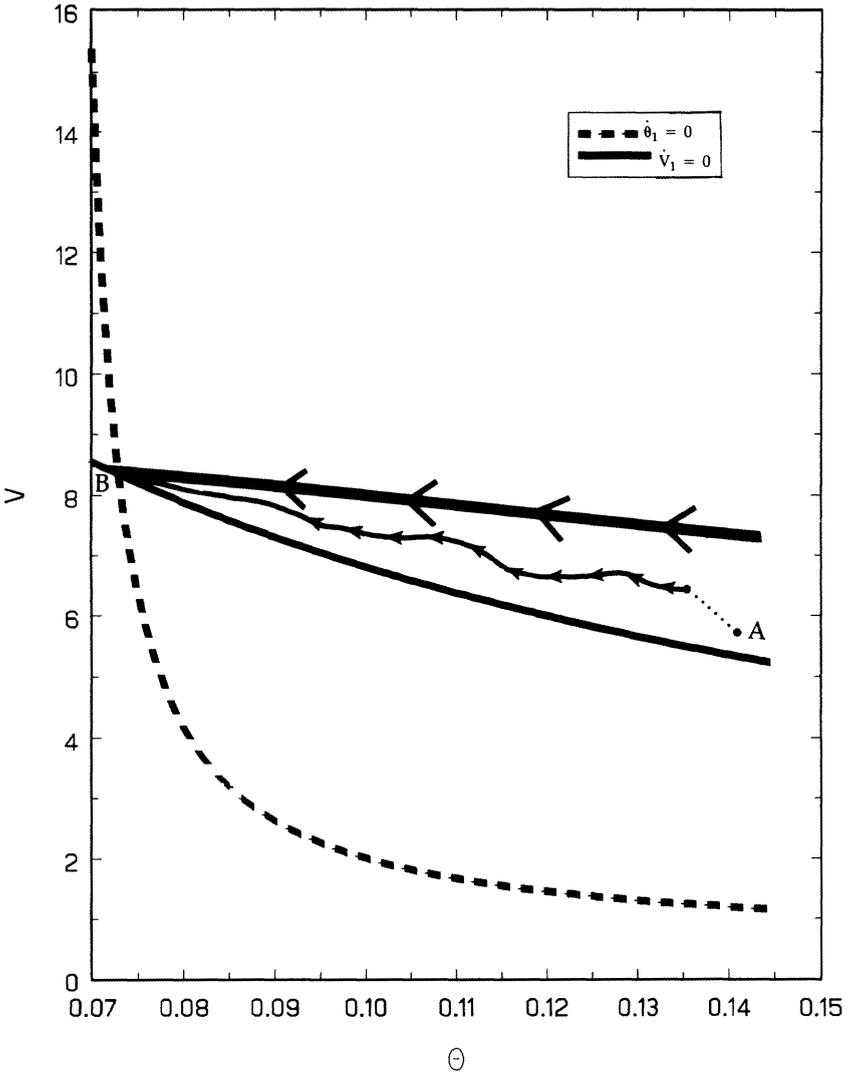


Figure 11: PHASE DIAGRAM



noninstantaneous diffusion. Point A corresponds to a steady-state equilibrium with the parameter configuration of the 1960–1974 period described earlier, while point B illustrates the steady state emerging from the 1975–1989 period. The thick line with arrows illustrates the saddle path of the new equilibrium. Because in reality the shift in parameters may have been slow, and the decline in δ seems to have compromised only newer cohorts, it seems unreasonable to assume that the actual dynamics can be characterized in terms of the new saddle path. Instead, a path like the one depicted by the thin line with arrows seems more likely.⁶⁷

Figure 12 illustrates the long-run effect of changes in the speed of diffusion of ideas (γ) and in technological destructiveness (β), with the black dots representing the steady state of an economy with the parameter values we found for the United States. It shows that γ is large, in the sense that further increases in it do not increase equilibrium growth significantly. An increase in the destructiveness parameter β , by lowering the equilibrium productivity of labor in research, reduces equilibrium growth and raises the required value of a new idea.

Finally, we address the optimal subsidy issue, focusing on the case where $\gamma \rightarrow \infty$. We also assume that the subsidy to labor used in research is financed with a tax on labor used in production of consumption goods.

Setting $\bar{L} = \eta_t = 1$, and letting s be the subsidy rate (in terms of units of consumption), it can be shown that in equilibrium, \dot{N} is:

$$\dot{N} = \left(\frac{1 - \alpha}{1 - s} \right) \left(\frac{\delta}{\beta} - \rho \frac{(\alpha - s)}{\alpha} \right), \quad (24)$$

which is clearly maximized as $s \rightarrow 1$. As always, however, there is a tradeoff between long-run growth and current consumption. Indeed:

$$C_t = \left(\frac{1 - \alpha}{\alpha} \right)^{1/\alpha} \frac{\beta (\alpha - s)}{\delta (1 - s)} \left(\rho + \frac{\alpha \delta}{\beta (1 - \alpha)} \right) e^{N_t}, \quad (25)$$

which, for given N_t , is decreasing with respect to s , and reaches zero when $s = \alpha$. Because the utility function is logarithmic, the optimal subsidy rate must be less than α .

Because we have assumed that exogenous technological progress is

67. Note that the initial jump in θ is possible only if the initial change in δ involves the potency of older patents.

Figure 12a: \dot{N}

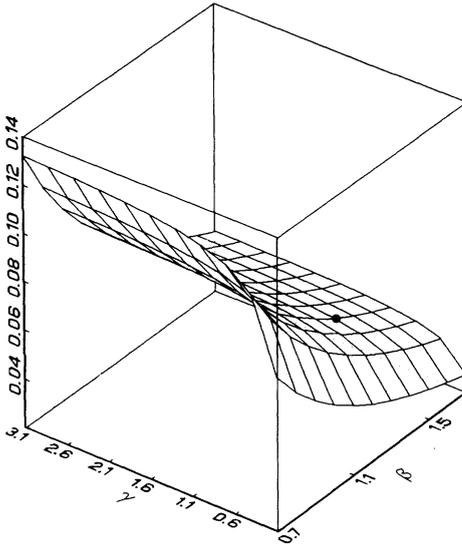
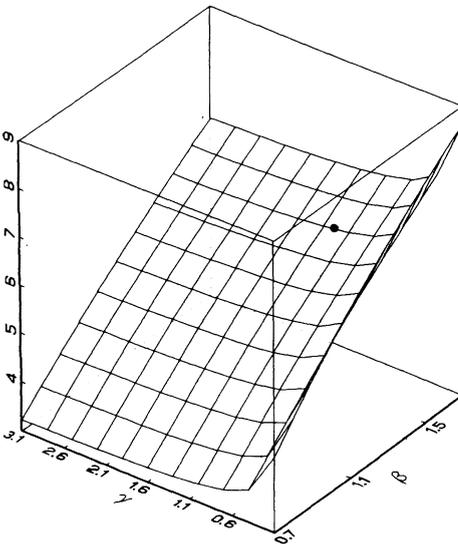


Figure 12b: V



negligible, we have that $\hat{C}_t = \dot{N}$, so we can write the present-value utility of the representative agent, U_0 , as:

$$U_0 = \frac{1}{\rho} \left\{ \ln C_0 + \frac{\dot{N}}{\rho} \right\}.$$

Maximizing this equation with respect to s , subject to Equations (24) and (25), yields the optimal subsidy rate, s^* :

$$s^* = \frac{\alpha^2(\delta/\beta - \rho)}{\rho + \alpha(\delta/\beta - 2\rho)}.$$

Replacing the parameters calibrated in the previous subsection yields an optimal subsidy rate of 33% if $\rho = 0.03$. If one turns back to Equation (24), such subsidy rate almost doubles the rate of growth of an unsubsidized economy characterized by the parameters calibrated for the United States.⁶⁸

5. Conclusion

We have constructed a model of economic growth through the creation of new goods, in which the phenomena of creative destruction and knowledge spillovers play prominent roles. The model has fairly simple and intuitive relationships between the existing public stock of knowledge and new ideas, between new ideas and growth, and between growth and the value of ideas or blueprints. The model produces endogenous growth for appropriate parameter values, and it highlights the importance of the speed of diffusion of existing knowledge and the endogenous rate of knowledge obsolescence.

We implemented the model empirically using patents as proxies for new ideas. First, we showed that it is possible to use patent citation information to put a fairly rich structure of knowledge diffusion and knowledge obsolescence onto the notion of research spillovers. We find that the rate of knowledge obsolescence rose from about 2 or 3% per year early in the century to about 10–12% per year at the end of the 1980s. Our results show that the process of knowledge diffusion is quite rapid, indeed, sufficiently rapid that the model performs essentially as if diffusion were instantaneous. In this context it is important to note

68. The optimal subsidy rate experiment raises the issue on whether our calibration exercise should be corrected to consider the fact that in the United States the subsidy rate is nonzero. We do not think that the precise numbers should be taken that literally.

that the lag we are measuring is between the grant date of the cited patent and the grant date of the citing patent. It seems plausible to view diffusion as beginning with the patent grant, because that is when the patent information is public. But the grant date of the citing patent is, of course, several months to a few years after its application date, and we take application date as being associated with invention. Thus, from the grant date of the cited patent to the application date of the citing patent would be even a shorter lag. Our results on the speed of diffusion seem to be broadly consistent with earlier work, particularly that of Mansfield (1985), who found that 70% of product innovations were known and understood by rivals within 12 months of the innovation, and only 17% took longer than 18 months.

This rapid diffusion rate prevented the large influx of foreign patenting in the United States in recent decades from lowering U.S. R&D productivity even further: With diffusion this rapid, the spillovers from the foreign knowledge creation approximately balance the increased rate of knowledge obsolescence that they also create.

This “good news” is overshadowed, however, by a measured reduction in the usefulness of existing public knowledge in generating new knowledge, as reflected in citation patterns. The estimated spillover potency (δ_s) fell by a factor of 5 over the century, with most of this occurring in the first few decades, and a fall of about 25% in the postwar period. When we translate this into the change in effective accumulated public knowledge, we predict a fall in the private productivity of research inputs of about 30% between the late 1950s and 1990.

We then move to the estimation of the innovation production function, the relationship between aggregate U.S. private research inputs and aggregate U.S. idea generation, as represented by U.S. patents. We confront the well-known “puzzle” of the large fall in the ratio of U.S. patents to U.S. research inputs in the postwar period. The citation function estimation could, potentially, explain this in two ways. If the size of patents was increasing fast enough, then the idea/research-input ratio may not be falling even if the patent/research-input ratio is. Second, if the effective stock of public knowledge is falling, then the reduced spillovers would explain the fall in the productivity of private research inputs. We find evidence of both effects, although the increase in patent size peaks in the early 1970s, so that our ideas/research-input ratio actually falls faster than the patent/research-input ratio after that. For the entire 1958–1990 period, we can explain the overall patent-productivity trend quite well, but we do not explain the accelerated decline in research productivity that occurred in the late 1970s, or the apparent reversal of the trend in the mid-1980s. One difficulty with understanding the very recent movements is that these patents have not had much

time to be cited, so our estimates of both δ_s and ψ_s are very imprecise for the late 1980s. Given the large increases in the number of patents in this period, it will be interesting to see how these patents fare as time goes by.

As noted, we also found evidence that the “size” of patents has grown over the century, increasing by a factor of 3 from 1900 until 1940, and then by an additional 20% until it peaked in about 1970. This is consistent with previous conjectures about changes in the propensity to patent. The early rise, in particular, is probably traceable to changes in the legal treatment of patents and the “corporatization” of research (see Schmookler, 1966). It may also be that innovation has become more “systems” oriented as it has become increasingly science-based, so that each “invention” is actually a larger and larger package of component ideas. It is also interesting that we find the size of patents to be falling in recent years. There were two major institutional changes in the 1980s that might have been expected to affect the propensity to patent, in opposite directions. First, patent application fees were increased, and fees for patent renewal were instituted for the first time in the United States in 1981. These changes should have operated to increase the threshold for inventors to decide to make a patent application, reducing the propensity to patent. At approximately the same time, there has been a perceived increase in the strength of patent enforcement in the United States. This makes patents more valuable and should thereby increase the propensity to patent. Our results suggest that the latter effect may be empirically more important.⁶⁹

Next we looked at the relationship between the rate of idea creation and consumption or productivity growth. We showed that, after removing high-frequency movements, the growth rates of either consumption or labor productivity display movements over the last several decades that correlate quite closely with the rate of invention that we measured. Thus, in our model, the productivity slowdown—the long fall in the smoothed growth rate of productivity from the mid-1960s—can be traced back to a fall in the rate of new product creation, which itself can be traced to a fall in research productivity connected to a decrease in the potency of old knowledge in generating new ideas.

The coincidence in timing of the fall in patenting in the 1970s and the slowdown in aggregate productivity has been noted by others. We have a story consistent with those facts, but we cannot push it too hard because so many of our assumptions about lags between observables and unobservables cannot be tested.

69. As can be seen from Figure 6, there has been a large increase in the patent/research ratio in the late 1980s. This would also suggest a possible rise in the propensity to patent (fall in the size of patents).

Our final empirical innovation is the measurement of rates of creative destruction, using data on patents and value at the firm and sectoral level. Unfortunately, these estimates can only be made for a shorter time period in the 1960s and 1970s, because the construction of patent totals for these firms in the 1980s has not been carried out. This exercise does give reasonable estimates for many sectors, varying between 0 for Petroleum Refining and 25% per year for Drugs, with a mean of about 3.5% per year. The estimated time path of the average rate of creative destruction is somewhat surprising, falling from a high of 7% in the mid-1960s towards zero by 1981. A challenge for future work will be to try to find alternative data series that would permit a richer analysis of rates of creative destruction by sector and over time.

We then took the empirical parameter estimates back to the model and showed that the observed decline in the productivity of research has implications for the innovation rate, the growth rate, and the value of new ideas that are all roughly borne out. The model simulation also emphasizes the importance of the apparently rapid diffusion rate of knowledge. The fact that knowledge diffuses rapidly prevented what could otherwise have been an even greater productivity slowdown in the 1970s and early 1980s.

Stepping back from particular parameter estimates and the consistency of particular model blocks with observed trends, we have suggested an organizing framework for empirical research on the contribution of industrial innovation to aggregate growth. We believe that this framework offers many avenues for fruitful future work. Having demonstrated that the citation function works reasonably well, it would be interesting to go back to it and focus in more detail on issues of stochastic structure and identification. Further, to really understand the significance and interpretation of the observed decline in spillover potency, we need to look at the variations across sectors and geographic space in the size of patents, and in the diffusion and obsolescence rates. In principle, one could categorize citing patents by technological sector and by the national origin or U.S. state of origination. This would allow one to put a finer structure on our homogenous, public good called knowledge, examining, e.g., whether foreigners are slower to pick up knowledge in U.S. patents than are Americans. One could also, to some extent, examine whether knowledge seems to have a private component, by looking at whether the firm cites its own patents more often or more rapidly than it does patents owned by other firms.⁷⁰

Consideration of cross-country citation patterns suggests that more

70. There is evidence, e.g., that such "self-citations" are more prevalent for private firms than for universities, and that they come sooner in time than non-self-citations. See Trajtenberg, Henderson, and Jaffe (1992).

thought needs to be given to how to think about the rate of invention, the rate of consumption growth, and the stock of public knowledge in an open economy. We have modeled U.S. consumption growth as depending on U.S. invention, U.S. invention as depending on U.S. research, but the “public” stock of knowledge available to U.S. researchers as being the worldwide stock. With respect to each of these, our assumption seems superior to the alternative polar extreme, but reality is probably somewhere between the extremes.

An interpretation of the decline in δ_s is that research is steadily becoming “narrower” and, hence, generates fewer spillovers because each new idea is relevant to a smaller and smaller set of technological concerns. Empirical testing of this notion would necessitate incorporating multiple dimensions of product quality into the model, so that there would be a notion of “technological distance” between different inventions.⁷¹ This could perhaps be implemented empirically using the patent classification information,⁷² although the classification information is not available in computerized form for patents before the late 1960s.

Finally, it would be interesting to look at the connections among the private value of particular inventions, the creative destruction they produce, and the knowledge spillovers they generate. To some extent, one would expect that important patents would be high on each of these scales, but ideas also probably vary in the magnitude of both the negative and positive externalities they generate.

REFERENCES

- Aghion, P., and P. Howitt (1992). A model of growth through creative destruction. *Econometrica* 60:323–351.
- Bound, J., et al. (1984). Who does R&D and who patents. In *R&D, patents and productivity*, Z. Griliches (ed.), Chicago: University of Chicago Press.
- Cockburn, Iain, and Z. Griliches (1988). Industry effects and appropriability measures in the stock market’s valuation of R&D and patents. *American Economic Review* 78:420–423.
- Cohen, W. M., and S. Klepper (1992). A reprise of size and R&D. Mimeo, November.
- Evenson, R. (1991). Patent data by industry: Evidence for invention potential exhaustion? In *Technology and Productivity, the Challenge for Economic Policy*. Paris: OECD.
- Gort, M., and S. Klepper (1982). Time paths in the diffusion of product innovations. *The Economic Journal* 92:630–653.
- Griliches, Zvi (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28:291–330.

71. Ariel Pakes emphasizes this point in his discussion of this paper.

72. A version of this is presented in Trajtenberg, Henderson, and Jaffe (1992).

- Griliches, Zvi (1989). Patents: Recent trends and puzzles. *Brookings Papers on Economic Activity, Microeconomics* 291–330.
- Griliches, Zvi, ed. (1984). *R&D, patents and productivity*. Chicago: University of Chicago Press.
- Griliches, Z. (1979). Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics* 10(1):92–116.
- Grossman, Gene M., and Elhanan Helpman (1991a). *Innovation and Growth in the Global Economy*, Cambridge: The MIT Press.
- Grossman, Gene M., and Elhanan Helpman (1991b). Quality ladders in the theory of growth. *Quarterly Journal of Economics* 106:557–586.
- Hall, B., et al. (1988). The R&D master file. NBER Technical working paper No. 72.
- Hall, B., Z. Griliches, and J. Hausman (1986). Patents and R&D: Is there a lag? *International Economic Review* 27:265–283.
- Hausman, J., B. Hall, and Z. Griliches (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52:909–938.
- Jaffe, A. (1986). Technological opportunity and spillover of R&D: Evidence from firms' patents, profits, and market value. *American Economic Review* 76:984–1001.
- Jaffe, A., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, forthcoming (NBER working paper No. 3993).
- Kortum, S. (1993). Equilibrium R&D and the decline in the patent-R&D ratio: U.S. evidence. *American Economic Review: Papers and Proceedings*, forthcoming.
- Jones, C. (1992). R&D-based models of economic growth. MIT mimeo, November.
- Levin, R., et al. (1987). Appropriating the returns from industrial research and development. *Brookings Papers on Economic Activity* 3:784–829.
- Mansfield, E. (1985). How rapidly does new industrial technology leak out? *The Journal of Industrial Economics* 34-2, December.
- Mansfield, E., M. Schwartz, and S. Wagner (1981). Imitation costs and patents: An empirical study. *The Economic Journal* 91:907–918.
- Mansfield, E., et al. (1977). Social and private rates of return from industrial innovation. *Quarterly Journal of Economics* 91(2):221–240.
- Merton, R. K. (1965). *On the Shoulders of Giants*. New York.
- Pakes, A. (1986). Patenting as options: Some estimates of the value of holding European patent stocks. *Econometrica* 54:766–784.
- Pakes, A. (1985). On patents, R&D and the stock market rate of return. *Journal of Political Economy* 95:390–409.
- Pakes, A., and Z. Griliches (1984). Patents and R&D at the firm level: A first look. In *R&D, patents and productivity*. Z. Griliches, ed. Chicago: University of Chicago Press.
- Pakes, A., and M. Simpson (1989). Patent renewal data. *Brookings Papers on Economic Activity, Microeconomics*.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98:S71–S102.
- Segerstrom, Paul S. (1991). Innovation, imitation and economic growth. *Journal of Political Economy*.
- Schankerman, M., and A. Pakes (1986). Estimates of the value of patent rights in European countries during the post-1950 period. *Economic Journal* 96: 1077–1083.

- Schmookler, J. (1966). *Invention and Economic Growth*. Cambridge, MA: Harvard University Press.
- Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. New York: Harper.
- Stokey, N. L. (1992). R&D and economic growth. Mimeo, June.
- Trajtenberg, M., R. Henderson, and A. Jaffe (1992). Ivory tower versus corporate lab: An empirical study of basic research and appropriability. NBER working paper No. 4146, August.

Comment

PHILIPPE AGHION
Oxford University and E.B.R.D.

This paper is an important contribution to the so-called neo-Schumpeterian theory of economic growth. The basic trade-off emphasized by the Schumpeterian approach is that between the positive spillovers induced by current innovations on future research activities on the one hand, and the creative destruction (or "obsolescence") effect exerted by new technologies on the existing ones on the other hand. The interplay between the positive and negative externalities respectively involved in the previous two effects is known to have far-reaching implications for the growth process, both positive and normative.

This paper contributes both to the theoretical and empirical understanding of the previous trade-off. It improves upon existing models of creative destruction, first by relaxing the extreme assumption that successful innovators become monopolies, and by allowing instead for (asymmetric) monopolistic competition between old and new technologies: Instead of being instantaneously displaced from a full monopoly position by new technologies, the old technologies can rely on the existence of product differentiation in order to remain in existence (although with a reduced market share) even after newer technologies have been implemented. An interesting consequence of mixing up quality ladders and product variety models is that the variety parameter (which also measures the degree of product market competition) affects both current profit and the equilibrium amount of creative destruction: The higher the degree of substitutability between older and newer products, the larger is the creative destruction affect.

The second, and perhaps more important, theoretical contribution of this paper lies in the endogenization of the research technology. More specifically, the productivity of research now depends in a sophisticated way upon the whole technological history, i.e., upon the time distribution of past innovations. This, in turn, can help us understand why the

productivity of research activities may evolve over time and in particular why it has decreased in the United States in the recent decades.

More precisely, this paper provides the *first* empirical analysis of research spillovers and creative destruction based on *microeconomic* patent data. Then key insight is that data on patents *and* on their citations by subsequent patents can serve as a basis to estimate both the degree of intertemporal spillovers (i.e., the positive externality that any innovation exerts on the subsequent ones) and the speed of knowledge obsolescence (measured by the speed at which the number of citations of a given patent by subsequent ones declines over time). The main result from this empirical analysis is that research spillovers generated by successive cohorts of patented innovations, and measured by their subsequent citations by future patents, have substantially declined over this century, hence pointing to a decline in the productivity of research. (This in turn might account—although the authors remain very cautious on this point—for the slowdown in aggregate productivity recently observed in most developed countries.)

Two main explanations are suggested by the authors in order to account for the decline in the average number of citations by future patents. One is the fact that the size of patents, measured by the number of new ideas embodied in them, has substantially increased over the century. In particular, the number of future patents that could potentially cite current or past innovations tends to decrease over time. This first explanation only accounts for the decreasing rate of citations, but not for the decreasing research (and output) productivity.

In this respect the model may not appear entirely consistent in the sense that the size of innovations, as measured by the quality parameter q , is supposed to be constant over time, whereas the citation rate, which also reflects the size of an innovation, is allowed to vary over time. Endogenizing the parameter q might in turn destroy the linear relationship between \dot{N} and C and, therefore, the interpretation of the results in terms of aggregate productivity slowdown.

On the other hand, there may be strategic and/or organizational reasons for why the rate of patent citations has declined over time. For example, large firms (i.e., firms with a long purse) often generate or purchase patents that they do not use (or the use of which they decide to delay), primarily in order to prevent entry by potential competitors on the same product market. This, however, would not account for the alleged slowdown in the productivity of research, even though it would be consistent with the observed decline in the average rate of citations.

More generally, identifying patents as a reliable measure of both the production and diffusion of new ideas and technologies independently

of the organizational environment may be quite misleading. In particular, it is known that research activities are increasingly governed by complicated inter- or intrafirm contracts whereby property rights on innovations are often split between research employees (or research units) and their employers (or customers), and where the governance structure (full vertical integration, research joint venture, etc.) has an important effect on the incentive to produce and/or diffuse new technologies. The relationship between new ideas and patents is thus bound to evolve over time and follow the evolution of the organizational and legal frameworks for R&D activities: In other words, patents may just be the tip of a more complicated iceberg, with no invariant relationship between the rate of patents citations and the rate of knowledge obsolescence.

A second explanation for (or possible implication of) the observed decline in the rate of patent citation in the United States is the increasing rate of knowledge obsolescence. Again, the relationship between these two phenomena is far from obvious. First, as the rate of technological discovery has greatly accelerated in recent decades, all the parties involved in patent citation (the innovators and his or her examiners) may share the same incentive to avoid informational and administrative overload and, therefore, reduce the number of patent citations prior to a given time period. In that case, a declining rate of patent citation would be no evidence either of an increasing rate of knowledge obsolescence or of a decline in the productivity of research.

Second, as suggested in the paper, innovations may have become increasingly specialized. This would instead be consistent with an increasing rate of obsolescence although not necessarily with a decline in the productivity of research. In order to be further investigated and tested, this latter explanation would require that the basic neo-Schumpeterian model be extended so as to introduce the distinction between fundamental (or general) and secondary (or specialized) innovations. This in turn would allow for the possibility of Schumpeterian waves in equilibrium and for a richer discussion of the dynamic evolution of technological spillovers and obsolescence. In particular, it might be interesting to investigate whether the alleged decline in the productivity of research over the last (two) decades is bound to be infinitely durable or instead corresponds to the temporary downfall part of a Schumpeterian wave.

Comment

ARIEL PAKES

Yale University and NBER

This is an ambitious and creative paper. It tries to mix theory with data to throw some light on how the interactions among agents in the production of knowledge affect growth and welfare. The mix of theory and data serves to focus discussion on magnitudes that might have both empirical and theoretical content. It also generates a framework that allows us to use the empirical estimates to evaluate various policy and descriptive scenarios. The authors view their framework as a structure that will evolve over time as they interact more with different types of data. My comments will be made with subsequent reformulations of their framework in mind.

1. *The Model*

The theory is very simple, as it should be in this stage of research. Products are differentiated by a “vertical”-quality dimension. New products are produced by research resources, have the same cost of production as old products, and extend the quality frontier. The productivity of research in producing new products depends on the information made available by the outcomes of past research. (Past research generates a nonpecuniary externality to new research.) Given the products that have been developed, a Nash equilibrium in prices determines current returns. That equilibrium implies that new products “obsolete” old products. (New products cause old products to earn less; a pecuniary externality.) Note that there is *no* physical obsolescence of products; if there are no new products then the value of the knowledge embodied in the old products does not decay over time. Although this is not correct for all types of knowledge, it is probably close enough to correct for the vast majority of the output of industrial knowledge-producing activities. (An example of a type of knowledge whose physical productivity does decay through use is the knowledge embodied in insecticides; frequently use of an insecticide leads to the development of resistant strains in the insect population.) There is a single consumer who consumes all products and generates the needed demand curves, while a free-entry assumption together with equilibrium in the labor market determines the amount of product development.

Note that there is only one dimension to the quality of a product.

This simplification should only bother us if it constrains the empirical work in ways that are either at odds with, or causes us to misinterpret, the data. I will note later where this assumption may become problematic. On the other hand, it is not obvious that we can do away with the assumption that quality has only one dimension without a much more extensive modeling effort. They make a second simplification, however, which is probably both more problematic and easier to rectify.

They currently do not allow for either any uncertainty or for any unobserved factors to affect the relationship between research inputs and research output. One of the striking results of empirical work in this field is precisely the enormous variance in the outcomes of industrial research activity, and there is good reason to think that this variance does not "average out" when we look at the sum of individual outcomes. In the more directed research that policymakers are usually concerned with (e.g., energy saving research or research on reducing emissions of pollutants), the fact that the randomness does not average out across research enterprises is because the different firms are all exploring the same unknown frontier, and the extent to which that frontier can yield to current knowledge is a common factor that affects the outcomes of all research. Further, there is a good deal of evidence that shows that the productivity in research in different fields is highly correlated. That is, historians tell us that there are periods of relatively rapid technological development, usually foreshadowed by an advance in a "metatechnology" that can be used in the development of many new types of products (synthetic materials in the postwar period, and semiconductors thereafter; see the classic article by Rosenberg, 1974).

I worry about ignoring the unobserved components in the relationship between industrial research inputs and outputs both because I think we could incorporate it in ways that would enhance the framework significantly, and because the exclusion of it is going to generate biases in the rest of our estimates. First on what we miss. In reality the productivity of industrial research depends on the outcomes of the (more basic) research done in our universities and national research labs. It would be interesting to hook the unobserved component of the productivity term back to this, more basic, research and quantify its impacts. In this context, I should note that patents not only cite other patents, they also cite key articles that were used in their development. By ignoring the unobserved productivity components, we are also giving up on the problem of coming up with a quantitative measure of the uncertainty associated with research outcomes. This can be important in its own right, especially in more detailed models where the extent of

uncertainty might have an impact on both how much and on which type of research we want to get engaged in.

The estimation problem created by ignoring the unobserved components of the outcomes of the research process is just the traditional simultaneous equations problem in estimating production type relationships. (The classic reference here is Marschak and Andrews, 1944.) Periods when the unobserved productivity is high will be periods when firms are induced to expand their research activities, so standard least-square analysis of measures of research output against research input will give you biased coefficients. (For a recent attempt to handle this problem at an industry level of aggregation, see Kortum, 1993.)

2. *Overview of the Empirical Analysis*

The empirical analysis has four equations:

1. A citation equation that is used to obtain both a measure of the extent to which past research has an impact on current research productivity (the nonpecuniary externality discussed earlier), *and* the relationship between patents and knowledge output.
2. An innovation production function in which “inputs” are current research expenditures and the citation measure of current research productivity, and the output is the patents-based measure obtained from 1.
3. A growth equation that relates either consumption or labor productivity to the measure of the outcomes of research activity.
4. A creative destruction equation that analyzes the relationship of stock market values of firms to a firm’s own patents, and to the patents of the other firms in the same industry. (The latter gives us the effect of the pecuniary externality discussed earlier.)

I begin with some general points and then focus in on a more detailed discussion of the first equations in this system.

Although much of the modern growth literature is focused on the impacts of externalities of one form or the other, almost all of the empirical work that emanated from it used only aggregate data. The essence of externalities in the interrelationships among units, and the externality impact we are looking for is often a difference between a “micro” and an “aggregate” response function. It is hard to believe that one could provide convincing evidence of the impact of interrelationships among agents, and/or of a difference between aggregate and micro response

functions, from an empirical analysis that relied solely on aggregate data. This paper actually goes back to the micro fundamentals and tries to estimate the impact of the externality generating relationships directly. This is a much more convincing way of determining the importance of the basic assumptions underlying the model, and it is also likely to be much more fruitful in pointing to directions where both our models and our empirical work are likely to bear fruit.

The second striking aspect of the empirical work is the use of citation data. To measure externalities effectively we need to know which bit of knowledge is "closer" to which other. (If they all have the same impact, then there is only an aggregate effect that is the same to all and, hence, whose impact is very hard to distinguish from the impact of other time invariant variables.) The citation variable is an independent measure of the "distance" between different "bits" of knowledge and, hence, can be used to examine the quantitative impact of various forms of externalities. (And recall that it is these externalities that lie at the heart of any difference between the social and private return to research and, hence, of policy analysis.)

The citation data are used here to obtain a measure of the impact past knowledge has on the production of new knowledge (the nonpecuniary externality noted earlier). Here the basic concept is that the patentees cite patents whose information content was used in the production of the knowledge embodied in the patent. This, however, is not the only possible use of the citation data. Because the purpose of the patent references is to cite prior state of the art, one could have thought of using citations to construct a measure of the extent to which different pieces of knowledge have been made obsolete. This would have allowed us to use the citation data to help construct measures of obsolescence for the analysis of creative destruction. More to the overall point, there is probably some content to both of these interpretations of the citation data, and, had we done both types of analysis, we might have learned more about how best to use the citation figures in future research.

3. The Citation Equation

This equation serves two purposes:

1. It determines the usefulness of past cohorts of patents in current knowledge production and, hence, determines current research productivity.
2. It determines the relationship between knowledge and patents.

To see how both these things are done in one equation, it is helpful to examine the structure of that equation. Its dependent variable is a_{ts} = cites to cohort s from patents sampled in t per patent in t divided by number of patents in s . If we look at their specification in logs and do not impose some of their constraints on functional forms, what we have is

$$\ln(a_{ts}) = g_1(t) + g_2(s) + g_3(t - s) + g_4\left(\sum_{j=s}^t \Delta \text{knowledge}_j\right),$$

$$\Delta \text{knowledge}_j \equiv \Delta N_j = \psi_j P_j,$$

where P_j is patents granted in j , and ψ_j measures the knowledge per patent in cohort j .

Here $g_4(\cdot)$ measures the degree to which patents in cohort “ s ” have become obsolete and, hence, contain less useful ideas, for cohort “ t ”’s use. In a unidimensional quality mode, where each good improves on the quality of other goods, it makes some sense to make the extent of obsolescence of period s goods a function of the number of new goods, or the amount of “new knowledge,” produced between s and t , and this is what they have done. Note, however, that this determines the relationship between patents and knowledge for the rest of the analysis, i.e., knowledge per patent in a cohort becomes equal to the obsolescence effect per patent of a cohort.

In a model with more than one quality dimension, this would no longer seem like a natural assumption, and, indeed, it is easy to think of new patents that actually enhanced the productivity of the ideas in old patents. So I would have preferred to have a different weight for patents in $g_4(\cdot)$ than the weight given to patents in the definition of ΔN . This would have required more econometrics, probably a simultaneous equations analysis of the whole system, and as a result may best be left for a different analysis. Still this is one case where the added econometrics might well be worth it; the measure of research productivity and the measure of knowledge increments are the basic inputs into the subsequent analysis of how knowledge affects the economic magnitudes in the subsequent analysis.

The other factors affecting a_{ts} are as follows:

$g_1(\cdot)$, the time effect, is meant to pick up changes in the institutional regime that lead to more citations for a given number of used ideas. (This appears in the paper as $\psi_t \phi_t$).

$g_2(\cdot)$ provides a measure of the importance of the ideas in a cohort for subsequent citations; i.e., it is our measure of the “externality intensity” of a cohort. (This appears in the paper as $\psi_s \delta_s$).

$g_3(\cdot)$ is a function of age that they argue has to do with diffusion of the ideas in a cohort [appears in the paper as a transform of $\gamma(t-s)$].

It is well known that one cannot estimate free functions of cohort, age, and time, so they have to impose some constraints. Their specification allows $g_1(\cdot)$ to be free (and it picks up a secular increase in citation intensity), approximates $g_2(\cdot)$ with a combination of dummies and polynomials, and makes linearity assumptions on both $g_3(\cdot)$ and $g_4(\cdot)$.

I pause here to consider the implications of the specification for $g_4(\cdot)$, which has

$$g_4(\cdot) = \beta \sum_{j=s}^t \psi_j P_{t-j},$$

with the normalization $\beta = 1$. The linearity assumption is a very strong assumption, which is not really consistent with past micro work of Jaffe and some of his coauthors. It bothers me for a couple of reasons.

First, one of their major findings is that there is a secular decline in the $\{\delta_s\}$, or in the usefulness of cohorts of patents in producing new knowledge. This is a major part of their explanation of the secular decline in the patent to research ratio. I am worried that the estimated fall in the $\{\delta_s\}$ series is simply an artifact of the fact that old patents are cited more than their exponential decay warrants, i.e., if we had a longer time series, the current generation of young patents would also pick up larger δ_s 's.

More generally the linearity in $g_4(\cdot)$ is also determining the shape of the $\{\psi_s\}$, which in turn is fed back into their analysis of the productivity of research resources and, hence, affects all subsequent analysis. As a first step to seeing whether this linearity is indeed a problem, they might have used some more flexible functional forms here (add a quadratic and maybe a cubic term, or, better yet, try a combination of dummies and polynomials that would allowed for bell-shaped patterns). Here are some more general comments on the empirical results from estimating the citation equation.

The most important thing to take from Table 2 is that allowing for intervening ideas (in the form of patents) in addition to the time and cohort effects, improves the fit quite markedly. This makes it clear that there really is empirical content to their notion of a “knowledge clock,”

the notion of obsolescence they are after. My problems, then, are just in the detailed interpretation they provide for an effect that is clearly present. Their interpretation left me troubled in the following ways.

First, they find that the $\psi_j = g_1(j)$ are almost perfectly negatively correlated with the $\delta_j = g_2(j)$. If I interpret this correctly, it says that cohorts that are cited disproportionately actually do not make obsolete as much past old knowledge as cohorts that are not cited as much. This may well be true, but my intuition says that if it is, then the model should have more than one “quality” dimension, in which case there is something wrong with the underlying framework we are using to interpret the relationships in the data. In particular, how do we know that it is the ψ_j that should be used to construct the $\Delta N'$ in the subsequent analysis, and not some combination of the ψ_j and the δ_j (particularly because firms tend to have base patents that they build on and cite with their subsequent work)?

Second renewal data indicates a negative correlation between quantity and the average *private* value of the patent rights in a cohort of patents. (Empirically this is just a negative correlation between the size of the cohort and the fraction of the patents in the cohort that find it profitable to pay renewal fees to keep their patents in force in subsequent years.) The simplest explanation is that more stringent rules at the patent office cause the less valuable patents not to apply, and because the patent distribution is highly skewed with many patents of very low value, the induced change in the number of patents in a cohort can be quite large even though the change in the total value of the cohort is negligible. (See Pakes and Simpson, 1989, and the literature cited therein.) Renewal data attempt to measure the private value of patent rights, and this value is different from the value of “patents” in augmenting aggregate productivity or aggregate consumption (which is closer to a social, than a private, value concept). Still we might have expected the two concepts to be positively correlated, and it bothers me that this paper finds, for the most part, a positive correlation between their quantity and “average value” estimates of cohorts of patents.

Finally, the standard errors in this and subsequent tables are probably not believable. Part of the problem is that they never are explicit about where the disturbances enter. (In principle this should be either where the researchers think they are missing an important piece of data or where there is true randomness in outcomes.) As a result, it is hard to figure out appropriate ways of calculating standard errors, but what they actually do is not robust to much. For example, a reasonable place to think the model is missing is in using a constant cite intensity of a cohort, so say we change to $g_2(s) + \epsilon_{ts}$, and let ϵ_{ts} be serially correlated

over time. That would imply both dependence in the residuals of their equations, and heteroscedasticity. The current standard errors do not make a correction for either of these.

4. The Last Three Equations

As far as I know all the citation analysis that has appeared in the economic literature to date is, in a sense, self-contained, i.e., it uses citations to measure which patents are closer to which others, but it does not bring that measure of closeness to bear in the analysis of other economic phenomena we want to explain. The innovation production function in this paper does just that; it asks whether the citation based measure of research productivity (θ) can help explain the relationship between industrial research and patents, and this is an important next step in the analysis.

The estimation results indicate that θ does indeed help explain research productivity. Indeed, to see this we need only compare the series of the ratio of their patents-based knowledge measure to research expenditures ($\psi P/R$), to the series on θ . The troublesome aspect of their finding is that it all seems to be coming from two trends; one in $\psi P/R$, and one in θ , i.e., the higher-frequency movements of the two series do not seem to be correlated. Thus, I would have liked them to also have allowed a time interaction, and see if their results survive. If they don't, and my guess is they will not, then the results should be interpreted as providing one potential source of a time trend in the patents to R&D ratio; other heavily trending variables are other possibilities. So I view this as perhaps promising, but not strong evidence, in support of a hypothesis. Consequently they have not totally convinced me of the usefulness of citation data in determining the productivity of research resources in producing patents. That is another reason that I would have liked to see the citation data used in the obsolescence analysis.

I should note here there are data reasons why they might not be picking up the higher-frequency movements in the $\psi P/R$ series. First, they have done a serial correlation adjustment in the innovation production function analysis. This seems to me to produce an inconsistency in the timing used to construct θ , i.e., if patents are a distributed lag over past knowledge increments, then the citations of those patents should also be a distributed lag of the citations made by the past knowledge increments. Also they use patents by grant date in the first part of the analysis and patents by application date in this part of the analysis; and there is a long and variable lag between these two series. These kinds of problems put real noise into the data, and my guess is that they

would help their results a lot by correcting them. (It might, for example, allow them to pick up the higher frequency movements in the $\psi P/R$ series.)

Finally, two brief comments on the analysis of creative destruction. First, I reiterate that it would be interesting to integrate the citation information into the stock market analysis. Second, I worry that the current analysis of growth of firm value ignores the growth in the other assets of the firm (growth in physical capital, e.g.), especially because we have good reason to believe the growth in physical capital is related to the growth in the firm's new products. In this context, it might be easier to build a specific stochastic structure into the model and then estimate parameters off the one period rate of return on the firm's stock (which has a cleaner interpretation in terms of revisions of information).

To conclude, let me point out that despite all these caveats, the authors get interesting results on every equation. What I take from this is that there is a data base available that should allow us to fill in many of the details needed to use their framework to understand the growth process.

REFERENCES

- Marschak, J., and W. H. Andrews (1944). Random simultaneous equations and the theory of production. *Econometrica* 12:143–205.
- Pakes, A., and M. Simpson (1989). Patent renewal data. In *Brookings Papers on economic activity, microeconomics*, M. Bailey and C. Winston (eds.), pp. 331–401.
- Rosenberg, N. (1974). Science, invention and economic growth. *Economic Journal* 90–108.
- Kortum, S. (1993). Estimating an equilibrium model of R&D. Mimeo, Boston University.

Discussion

While lauding the results in the paper, Zvi Griliches suggested that the citation data must be interpreted carefully. For example, the fact that citations to a patent cease does not necessarily imply the knowledge associated with the patent has been made obsolete: Economists today readily use the Cobb-Douglas production function without citing Douglas explicitly. Moreover, as the authors are certainly aware, patents and citations are an artifact of a certain process, and that process is distortionary. Citations do not necessarily reflect the knowledge that influences the inventor; often they are placed into the article by the patent

editor. As another example, the body of knowledge available for citation is getting larger and larger, but the bibliography of a patent has a natural size constraint, i.e., how much the editors will accept. Patents that might otherwise be cited will be crowded out, but this need not reflect obsolescence.

Regarding the relatively quick diffusion found in the paper, Griliches noted that again the patenting process may be relevant. Patent examiners are responsible for specific types of patents. Therefore, when a patent application is received, the patent examiner is cognizant of the most recent patents granted in the field and may require these in the list of citations. The diffusion measured in the paper is the diffusion among patent examiners, not the diffusion of the innovation in the average productivity of the economy.

Roland Benabou wondered if there have been sudden changes in patent regulations such as application fees, royalty rates, or the duration of patents that might be used to distinguish variations in the propensity to patent from the other effects that we are really interested in. Jaffe responded that the largest change in the United States was in the early 1980s, when patent renewal fees were instituted. Beginning in 1985, data is available on whether or not inventors renew their patents. Pakes and others have worked with European patents where renewal fees have been in place for a longer period of time. Griliches noted that in recent years, patent fees have also increased sharply. Interestingly, Jaffe remarked, the rate of patenting has risen dramatically over this same period.

Benabou also suggested that royalty fees should be included in the valuation equation for patents: If your innovation will require paying royalty fees on the patents that are cited, this will presumably affect the valuation of the innovation. Jaffe responded that in fact, an inventor generally does not have to pay the owners of patents that are cited. This depends on the nature of the product that is manufactured and sold. If you sell a product that incorporates both your invention and the cited inventions, you have to pay royalties. In principle, however, a patent is a statement by the patent office that your invention is sufficiently freestanding that it is a product in and of itself, and no royalties are required.