INTERNATIONAL ECONOMIC REVIEW
Vol. 37, No. 4, November 1996

PATENT LENGTH AND THE RATE OF INNOVATION*

BY ANDREW W. HOROWITZ AND EDWIN L.-C. LAI[1]

This paper models the effect of patent length on the rate-of-innovation and consumer welfare. We find that the patent length that maximizes the rate-of-innovation exceeds that which maximizes consumer welfare. We show a countervailing effect of patent length upon the "size" and "frequency" of innovation. Longer patents increase the size, but decrease the frequency of innovation. The patent lengths that maximize the rate-of-innovation and welfare represent balance points between size and frequency. The divergence of the welfare maximizing and rate-of-innovation maximizing patent lengths has important policy implications that we briefly explore.

1. INTRODUCTION

In a sweeping review of U.S. technology and patent policy in 1988, the House Committee on Science and Technology repeatedly states the objective of increasing the "pace of technical change."[2] This emphasis on the *rate of innovation* departs from much of the theoretical patent literature, which focuses instead on the consumer surplus—deadweight loss tradeoff. Nordhaus (1969), in his seminal work on invention and welfare, sets the tone of this literature by developing a model in which patent length determines the "size" of innovation. The "optimal patent length" is that which induces an innovation of welfare-maximizing size. More recently, papers by Gilbert and Shapiro (1990), Klemperer (1990), and Gallini (1992), examine the interplay of the patent breadth and length instruments in determining optimal patent structure.

Another strand of the literature explores the effect of patent length in dynamic settings. Judd (1985) develops a general equilibrium model of continuous product innovation and finds that patents (or imitation lags) of infinite duration may achieve too much, too little, or the socially optimum level of innovation. He also demonstrates that finite-life patents may induce undamped oscillations in innovation. Jovanovic and Rob (1990) explore the link between discovery, imitation, and business cycles in a partial equilibrium model. They do not, however, discuss patent protection. Mookherjee and Ray (1991) develop a partial equilibrium model of successive innovation with focus on the timing of innovation as the rate of diffusion varies. Chou and Shy (1993) present a multiproduct, overlapping-generation model

* Manuscript submitted October 1994; revised May 1995.
[1] We would like to thank David Wildasin, E. Young Song, Robert Driskill, James Foster, Jennifer Reinganum, Sam Kortum, Robert Margo and anonymous referees for helpful comments. Remaining errors are ours.
[2] See House Report number 100-1093. The objective of increasing the pace of technical change is consistent with the stated objective of the first U.S. patent law (1790), to wit: "to promote the progress of science."

of product innovation in which the saving of the young can be allocated to innovation. However, only the polar cases of one-period and infinite patent length are considered, and the welfare-maximizing patent length is not identified.

While the analyses discussed above have provided important insights, they do not focus on the effect of patent length on the rate of innovation or *the relationship between the rate of innovation and consumer welfare*. This nexus, which was a central concern of Schumpeter (1939), is intertwined with the stated objective of patent law and is the primary focus of our analysis.

We model innovation as movement by a leading firm up a "product quality ladder." Higher positions on the ladder are associated with higher quality products. Imitating firms may follow the leader up the ladder, but are constrained in the dimension of time by patents.[3] We define the *size* of an innovation as the distance traveled on the product ladder and the *frequency* of innovation as the average number of innovations per-period. We define the *rate of innovation* as the average distance traveled up the product ladder per-period, which is the product of the size of innovation and the frequency of innovation. Many of the earlier models focused exclusively on the size of innovation, ignoring the effect of patents on innovation frequency.

Our model identifies a unique patent length that maximizes the rate of innovation. Patent length either too short, or too long, will weaken innovative incentives. We then derive the patent length that maximizes consumer welfare. We find that, in general, the patent length that maximizes the rate of innovation is longer than that which maximizes consumer welfare. This divergence creates a tension between the stated objective of patent law (to maximize the rate of innovation) and the more intuitive objective of maximizing consumer surplus.

The existence of finite patent lengths that maximize the rate of innovation and consumer welfare are attributable to the countervailing effects of patent length on the size and frequency of innovation. Longer patent length increases the size but decreases the frequency of innovation. For short patent length, the decrease in size of innovation dominates the increase in frequency, causing the rate of innovation to fall. For long patent length, the decrease in frequency dominates the increase in size, again resulting in a falling rate of innovation. At an intermediate patent length, these forces are balanced so as to maximize the rate of innovation.

The trade-off between size and frequency also generates the existence of a finite welfare-maximizing patent length. However, the decline in the frequency of innovation as patent life is extended affects consumers both directly (due to discounting) and indirectly (through its effect on the rate of innovation). Consequently, from a welfare perspective, the balance point between the size and frequency of innovation occurs at a shorter patent length. That is, the patent life that maximizes consumer welfare is shorter than that which maximizes the rate of innovation.

[3] In practice, patent protection depends on the time to circumvent the patent, as well as statutory length. If competitors' are unable to circumvent a patent before its expiration, statutory length is the time measure of protection. If circumvention occurs before expiration, then the time to circumvent is the appropriate measure of protection. In either case the strength of protection may be measured in time units as the interval between innovation and the erosion of the leader's market power.

The paper is organized as follows: Section 2A motivates the model. Section 2B formally specifies the consumer and firm problems and derives the preliminary results. Section 2C derives the relationship between patent length and the rate of innovation. Section 2D demonstrates the existence of patent lengths that maximize the rate-of-innovation and consumer welfare. Section 2E examines the relationship between the patent lengths that maximize consumer welfare and the rate-of-innovation. Section 3 concludes.

2. THE MODEL

A. Motivation. Our objective is construction of the simplest model capable of providing meaningful insight into the effect of patent length on the rate of innovation and welfare. We consider a product quality ladder with an innovating firm which obtains patents defined in the dimension of time as it climbs the ladder. Potential imitators stand ready to copy products whose patent protection has ceased. The presence of a single innovator approximates specialization of the innovative and imitative activities. Our quality ladder model is similar to that employed by Grossman and Helpman (1992; see also Aghion and Howitt 1992, and Segerstrom 1991). However, in contrast to the fixed rung property of the Grossman and Helpman ladder, we allow the innovator to traverse the ladder in arbitrary increments. That is, we employ a continuous ladder.

Products along the continuous ladder are perfect substitutes in the sense that there exists a quality-adjusted price at which consumers are indifferent between different positions on the ladder. Thus, though the leader adopts a limit-pricing strategy that prevents actual entry, the market is fully contestable. The threat of imitation disciplines the leader and induces behavior that approximates that under actual imitation.

B. Preliminaries. Consider a continuous product ladder with a single specialized innovative leader. The marginal cost of innovation, C, is assumed constant and is incurred in the dimension of distance traveled on the product ladder. Thus, the cost of traveling distance "d" up the product ladder is $C * d$. There exist potential imitators whose ability to produce along the product ladder is constrained in the dimension of time by a patent of length τ. That is, if the leader develops a product at time t, potential imitators cannot begin producing the product until time $t + \tau$. When the patent expires, we assume the cost of imitation is zero.[4] The unit cost of production (as distinct from innovation costs) along the product ladder are also constant at "η"per unit. We index time by $t = -1, 0, 1, 2, \ldots$, and designate $t = 0$ as the first period in which innovation can occur. The product existing at $t = -1$ represents a preexisting technology.

The demand side of the market consists of a representative consumer who maximizes utility in each period subject to a budget constraint. Let I_t be the set of

[4] This can be thought of as an approximation of an industry in which imitation costs are very low. For example, in the pharmaceutical industry products are often simply chemical compounds. Reproduction by other firms once the patent has expired is typically straightforward.

all products that exist at time t, where $i \in I_t$ indexes a particular good. Let $\upsilon(i)$ be the *position* of good i along the quality ladder. Consumer valuation of the *quality* of good i is represented by the function: $q(i) = \lambda^{\upsilon(i)}$, $(\lambda > 1)$. Without loss of generality we assume that the quality of the preexisting product is 1. The *quantity* of good i consumed at time t is denoted as $x(i,t)$. The consumer's problem is to choose a quantity (x), of each good i from the set of available goods at time t (I_t), subject to a budget constraint. That is,

$$\max_{\{x(i,t)\}_{i \in I_t}} U(t) = \sum_{i \in I_t} q(i) x(i,t) \tag{1}$$

$$\text{subject to } \sum_{i \in I_t} p(i,t) x(i,t) = E,$$

where $p(i,t)$ is the price of good i at t, and E is exogenously determined total consumer expenditure during each period.[5]

Firms engage in price competition each period. Since (1) implies that products of different qualities are perfect substitutes for each other and unit costs are constant along the product ladder, there exists a limit-pricing equilibrium in which only the highest-quality product will be produced. This equilibrium can be thought of as an approximation of one where entry actually occurs, in the sense that the leader is disciplined in price by the potential entrants.[6]

Let $n(t)$ be the *position* (along the quality ladder) of the highest-quality product existing at time t. That is, $n(t) = \max\{\upsilon(i) | i \in I_t\}$. Let $m(t)$ be the position (along the quality ladder) of the most advanced product that can be imitated at time t, given the patent law. Under limit pricing the leader, who produces only the highest-quality product, will set a price of

$$p_n = \eta \lambda^{n-m} - \varepsilon \qquad (\varepsilon > 0,\ \varepsilon \to 0) \tag{2}$$

where the time arguments are suppressed and p_n will henceforth refer to the price of the most advanced product available at a given time (recall that η is the unit cost of production). The lowest price at which the (potential) imitator will produce the good at position m is $p_m = \eta$, where (again suppressing the time argument) p_m is the price of the most advanced product that can be imitated. By adopting the limit-pricing strategy (2), the leader will capture the whole market, since the marginal utility per dollar for the good at position n exceeds that for the good at position m:

$$\frac{MU_n}{p_n} = \frac{\lambda^n}{\eta \lambda^{n-m} - \varepsilon} > \frac{\lambda^m}{\eta} = \frac{MU_m}{p_m}. \tag{2a}$$

[5] As with many patent models in the industrial organization literature, ours is a partial equilibrium analysis. Such an analysis allows us to solve for a steady-state rate of innovation for all patent lengths, and compare the patent length that maximizes the rate of innovation with that which maximizes welfare.

[6] This is essentially a contestable market argument.

Moreover, there is no incentive for the leader to price below $n\lambda^{n-m} - \varepsilon$ since it faces unitary price elasticity of demand below the limit price. Hence, there is a corner solution and specialization in consumption of the most advanced product. The net revenue (i.e., profit before innovation costs are subtracted) of the leading firm at time t is therefore

$$(3) \quad \pi_t = (p_n - \eta)x_n = \left(p_n - \frac{p_n}{\lambda^{n-m}}\right)x_n = \left(1 - \frac{1}{\lambda^{n-m}}\right)p_n x_n = \left(1 - \frac{1}{\lambda^{n-m}}\right)E$$

with $t = 0$ designated as the first period during which innovation occurs. After an innovation, the patent law dictates the position of the imitative threat. The complete description of the imitative threat is

$$(4a) \qquad m(t) = 0 \qquad \text{for } 0 \leq t < \tau$$

$$(4b) \qquad m(t) = n(t - \tau) \quad \text{for } t \geq \tau$$

Equation (4a) is simply an initial condition that says an imitator may not begin movement up the product ladder until the patent on the leader's first product expires at $t = \tau$.[7] Thereafter, the imitator's position on the product ladder at time $t \geq \tau$ is that which the leader occupied τ periods ago (i.e., $n(t - \tau)$).

Using equations (1) through (4) we can now write period t consumer utility compactly as follows:[8]

$$(5) \qquad U(t) = \begin{cases} \dfrac{E}{\eta} & \text{for } 0 \leq t < \tau \\[2ex] \dfrac{E}{\eta}\lambda^{n(t-\tau)} & \text{for } t \geq \tau \end{cases}$$

We now explore the effect of τ on firm behavior.

C. Patent Length and the Rate of Innovation. Returning to the firm's problem, the leader maximizes the present discounted value of the stream of profits given in (3) by choosing a position on the product ladder each time period. That is, the leader maximizes

$$(6) \qquad \max_{\langle n(t)\rangle_{\tau=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t[\pi_t(n(t), m(t)) - C[n(t) - n(t-1)]]$$

[7] Note that the initial condition must specify the position of the imitative threat from time 0 to $\tau - 1$.

[8] To see this first note that since only one product is consumed we can substitute (2) into the budget constraint of equation (1) to obtain an expression for $x(n, t)$. Then substitute $x(n, t)$ and $q(n) = \lambda^n$ into the objective function of (1). Recall that the quality of the preexisting product is 1. Then for $t < \tau$, invoking (4a) yields the upper part of (5). For $t \geq \tau$, invoking (4b) yields the lower part of (5).

subject to (4a) and (4b), where $0 < \delta < 1$ is the discount factor; C is the marginal cost of innovation, and we normalize by setting $n(-1) = 0$. Note that innovative cost is imposed in the dimension of distance traveled on the product ladder: $n(t) - n(t-1)$.

The first-order condition with respect to $n(t)$ is:[9]

$$\text{(7)} \qquad \delta^t \frac{\partial \pi_t}{\partial n(t)} + \delta^{t+\tau} \frac{\partial \pi_{t+\tau}}{\partial n(t)} - C\delta^t + C\delta^{t+1} = 0 \quad \text{for } t \geq 0$$

Using (3), (4a), and (4b), equation (7) becomes

$$\text{(8a)} \qquad [\lambda^{m(t)-n(t)} - \delta^\tau \lambda^{n(t)-n(t+\tau)}] E \ln \lambda = C[1-\delta] \quad \text{for } t \geq 0$$

The left side of (8a) is the marginal revenue of innovating the product $n(t)$, which is decreasing in $n(t)$. The right side reflects the marginal cost of innovation, which is constant. Rearranging equation (8a) and invoking (4b) we obtain:

$$\text{(8b)} \qquad \lambda^{n(t-\tau)-n(t)} = \frac{C(1-\delta)}{E \ln \lambda} + \delta^\tau \lambda^{n(t)-n(t+\tau)} \quad \text{for } t \geq \tau$$

Now by the same procedure, the first-order condition with respect to $n(t+\tau)$ can be written as:

$$\text{(8c)} \qquad \lambda^{n(t)-n(t+\tau)} = \frac{C(1-\delta)}{E \ln \lambda} + \delta^\tau \lambda^{n(t+\tau)-n(t+2\tau)} \quad \text{for } t \geq \tau$$

Substituting (8c) into (8b) we obtain

$$\text{(8d)} \qquad \lambda^{n(t-\tau)-n(t)} = \frac{C(1-\delta)}{E \ln \lambda}[1+\delta^\tau] + \delta^{2\tau} \lambda^{n(t+\tau)-n(t+2\tau)} \quad \text{for } t \geq \tau$$

Continuing recursive substitution of the first-order conditions with respect to $n(t+2\tau), \ldots, n(t+\phi\tau)$ yields

(8e)

$$\lambda^{n(t-\tau)-n(t)} = \frac{C(1-\delta)}{E \ln \lambda}(1 + \delta^\tau + \delta^{2\tau} + \cdots + \delta^{\phi\tau}) + \delta^{(\phi+1)\tau} \lambda^{n(t+\phi\tau)-n(t+(\phi+1)\tau)}$$

for $t \geq \tau$, and integer ϕ

[9] This procedure assumes perfect foresight on the part of the firm. To see that (6) is in fact concave, note that each row of the Hessian contains only two nonzero elements; one on the diagonal (which is negative) and one τ spaces to the right of the diagonal element. Using Laplace expansion, it is immediate that Hessian is negative-definite (see, for example, Chiang 1984, pp. 94–95). An implicit constraint, which is subsequently seen to be nonbinding, is that $n(t)$ cannot decrease over time.

Note that the last term on the right term of (8e) tends to zero when ϕ goes to infinity and (8e) becomes

$$\lambda^{n(t-\tau)-n(t)} = \frac{C(1-\delta)}{(1-\delta^{\tau})E\ln\lambda} \quad \text{for } t \geq \tau \tag{8f}$$

With the initial condition (4a), and the first-order condition (8a), equation (8f) indicates that the leading firm's optimal program consists of a series of discrete innovations of constant size at intervals of τ.[10] This behavior has a natural interpretation: the leading firm times the introduction of new products to coincide with the expiration of patents on earlier vintage products. The size of innovation undertaken each τ periods is:

$$n(t) - n(t-\tau) = \frac{1}{\ln\lambda}\ln\left[\frac{(1-\delta^{\tau})E\ln\lambda}{C(1-\delta)}\right]. \tag{8g}$$

Note that since δ is less than one, size is increasing in patent length. Letting $d(t)$ denote the size of innovation in period t, the complete innovation decision rule can be written as

$$d(t) = \begin{cases} \dfrac{1}{\ln\lambda}\ln\left[\dfrac{(1-\delta^{\tau})E\ln\lambda}{C(1-\delta)}\right] & \text{for } t = 0, \tau, 2\tau, 3\tau, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{8h}$$

Having established that an innovation of constant size occurs once every τ periods, the frequency of innovation is $1/\tau$. Since the rate-of-innovation is the product size and frequency, it is constant. Let z denote the rate of innovation. Then for given λ, δ, and C the rate of innovation is:

$$z = (n(t) - n(t-\tau))/\tau.$$

Substituting this expression into (8f) yields:

$$\lambda^{-z\tau} = \frac{C(1-\delta)}{(1-\delta^{\tau})E\ln\lambda} \tag{9}$$

The unique z which satisfies (9) given parameters τ, δ, λ, and C, is the profit-maximizing rate of innovation. Solving (9) for z we obtain

$$z = \ln\left[\frac{(1-\delta^{\tau})E\ln\lambda}{C(1-\delta)}\right]\frac{1}{\tau\ln\lambda} \tag{10}$$

[10] To see this note that (4a) and (8a) yield $\lambda^{-n(t)} = C[1-\delta]/\ln\lambda[1-\delta^{\tau}]$ for $0 \leq t \leq \tau$. Thus the leader innovates at $t = 0$, and remains at $n(t) = [1/\ln\lambda]\ln[\ln\lambda[1-\delta^{\tau}]/C[1-\delta]]$ for $0 \leq t < t$. Using (4b) and (8f), it is clear that the leader innovates again at $t = \tau, 2\tau, 3\tau, \ldots$. More generally, (8f) can be employed to completely characterize the time path of $n(t)$ for any initial conditions over the entire interval $0 \leq t \leq \tau$.

Equation (10) is the fundamental description of the rate of innovation in our model. Note that this rate of innovation is independent of initial conditions (equation 4a). Inspection of (10) reveals the following necessary condition for a positive rate of innovation

$$(1 - \delta^{\tau}) E \ln \lambda > C(1 - \delta) \tag{11}$$

If this inequality is reversed z will be negative, which we may interpret as a corner solution in which the rate of innovation is zero. Looking to the left side of (11), we see that consumers must have sufficient quality preferences (large enough λ), or the firm will have no incentive to innovate. The left side of (11) also reveals the existence of minimum patent length ($\tau^{\min}$), above which a perpetual sequence of discrete innovations will occur. Specifically, $\tau^{\min}$ is the smallest τ that satisfies inequality (11). It is immediate that in the absence of a patent law ($\tau = 0$), the rate of innovation will be zero. Inequality (11) also indicates the maximum marginal cost of innovation (in terms of the other parameters) at which the rate of innovation will be positive.

D. The Rate of Innovation and Welfare-Maximizing Patent Lengths. We have now obtained an expression relating the rate of innovation (equation 10) to patent length, and identified restrictions on parameters that ensure ongoing innovation. The question now arises; does there exist a patent length that maximizes the rate of innovation? Inspection of equation (10) reveals the countervailing effects of patent length (τ) upon the rate of innovation (z) that are necessary for the existence of an extreme point. An increase in τ decreases the frequency of innovation, which has a negative influence on the rate of innovation. On the other hand, a larger τ increases the *size* of innovation, which exerts a positive influence on the rate of innovation. The existence of a patent length that maximizes the rate of innovation requires a balancing of these two effects. Proposition 1 of the Appendix demonstrates that just such a balance point exists and that z is approximated by the form indicated by Figure 1. For ease of exposition we treat the patent length, τ, as a real number.[11]

We have now established the existence of a patent length that maximizes the rate of innovation. Is there a patent length that maximizes consumer welfare? If so, how does it compare to the patent length that maximizes the rate of innovation? To answer these questions, first recall from equation (5) that instantaneous consumer utility for time $s \geq \tau$ has the form $U(s) = (E/\eta)\lambda^{n(s-\tau)}$ and $U(s) = E/\eta$ for $0 \leq s < \tau$. Further recall that since innovation occurs every τ periods

$$\lambda^{n(j\tau)} = \lambda^{n(j\tau+1)} = \cdots = \lambda^{n((j+1)\tau-1)} = \left[\frac{(1 - \delta^{\tau}) E \ln \lambda}{C(1 - \delta)}\right]^{j+1}$$

[11] Since the time period is arbitrary, it can be made as short as desired. For example, time (and hence patent length) can be defined in units of months, hours, or even nanoseconds. As the units become shorter, our approximation becomes arbitrarily close to the integer mapping. Formally, we are examining an "extension" of z in which z is "embedded." See Saaty (1970, pp. 20–26) for a formal discussion of the use of extensions as approximations of integer mappings.

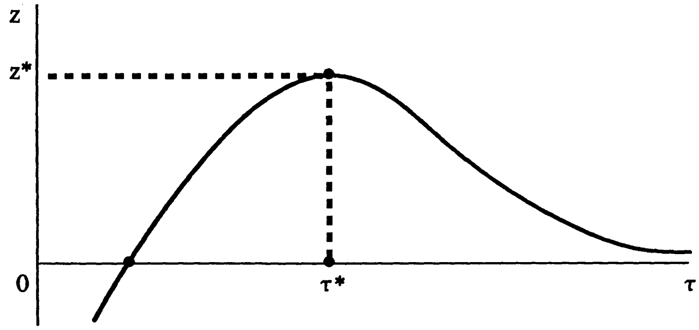

FIGURE 1

for all $j = 0, 1, \ldots, \infty$. Then assuming for simplicity that the consumer's subjective discount factor is the same as the discount factor employed by the firm (δ), intertemporal utility is

$$(12) \qquad w(\tau) = \sum_{s=0}^{\infty} U(s)\delta^s = \sum_{j=0}^{\infty} \sum_{s=j\tau}^{(j+1)\tau-1} \delta^s \left(\frac{E}{\eta}\right)\left[\frac{(1-\delta^\tau)E \ln \lambda}{C(1-\delta)}\right]^j.$$

With some manipulation it can be shown that

$$(13) \qquad \sum_{s=j\tau}^{(j+1)\tau-1} \delta^s = \left[\frac{1-\delta^\tau}{1-\delta}\right]\delta^{j\tau} \quad \forall j = 0, 1, 2, \ldots.$$

Since δ^s is the only term in which the time index "s" appears on the right-hand side of (12), we may substitute (13) into (12) to obtain the following expression for consumer welfare as a function of patent length:

$$(14) \qquad w(\tau) = \sum_{j=0}^{\infty} \left[\frac{1-\delta^\tau}{1-\delta}\right]\delta^{j\tau}\left(\frac{E}{\eta}\right)\left[\frac{(1-\delta^\tau)E \ln \lambda}{C(1-\delta)}\right]^j.$$

Note that the index "j" does not measure time. Rather "j" can be interpreted as denoting distinct vintages of product, each of which occupies the position of "most advanced" (and is consumed) for τ periods. With some manipulation (see Proposition 2 of the Appendix), equation (14) can be rewritten as:

$$(15) \qquad w(\tau) = \left[\frac{1-\delta^\tau}{1-\delta}\right]\left(\frac{E}{\eta}\right)\frac{1}{1-\left[\dfrac{(1-\delta^\tau)\delta^\tau E \ln \lambda}{C(1-\delta)}\right]}.$$

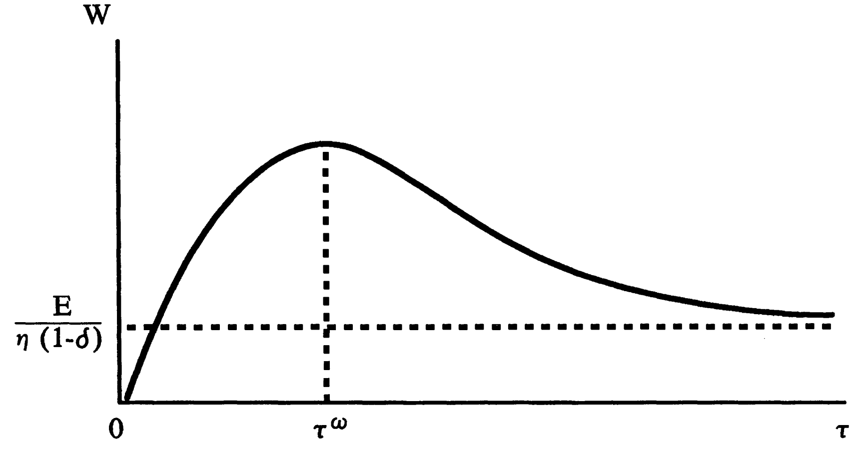

FIGURE 2

Proposition 2 of the Appendix also demonstrates that the graph of (15) over a real domain has the form illustrated in Figure 2 below.[12]

The intuition for the existence of a welfare-maximizing patent length is closely related to that of the rate-of-innovation maximizing patent length. As patent length is extended, consumer welfare increases so long as the increased size of innovation dominates the decreased frequency (from a welfare perspective). At the balance point of this trade-off, a welfare maximum is achieved. We formally analyze this trade-off in the following section.

E. Rate of Innovation and Welfare-Maximizing Patent Lengths: Relationship and Policy Implications. The natural question at this point is the relationship between the patent length that maximizes the rate of innovation (τ^*) and that which maximizes consumer welfare (τ^{ω}). To help answer this question, define $B = [[C(1-\delta)]/[E \ln \lambda]]^{1/2}$. Inequality (11) indicates that 1 is an upper bound on B. If $B > 1$, we have a corner solution with a rate of innovation of 0. Intuitively, at this corner either costs are too high or discount factor, income, or the taste parameters are too low to warrant innovation by the firm. In Proposition 3 of the Appendix we show that there also exists a lower bound on B of $1/2$. If $B < 1/2$ then w is infinite, and τ^{ω} is not defined. Proposition 3 of the Appendix establishes that for any parameter values for which τ^* and τ^{ω} are defined (i.e., $B \in (1/2, 1)$), $\tau^* > \tau^{\omega}$, as shown in Figure 3 below.

Figure 3 indicates that the patent length that maximizes the rate of innovation will, in general, exceed the length that maximizes consumer welfare.[13] The motiva-

[12] See footnote 11 for a justification of this approximation.

[13] The τ^* and τ^{ω} functions of Figure 3 are again real approximations of integer values.

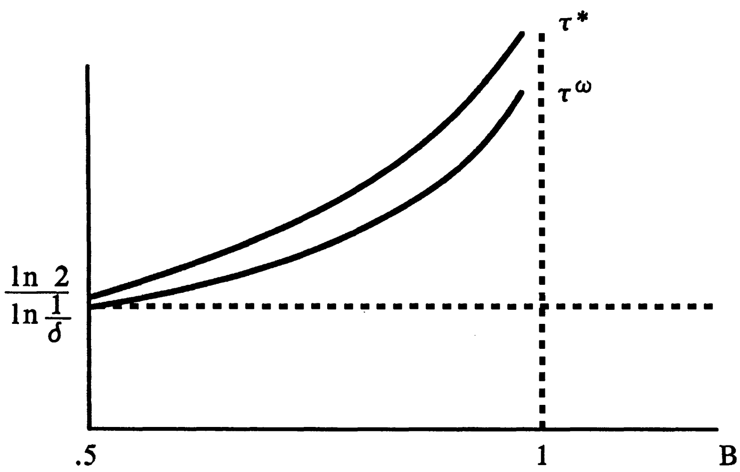

FIGURE 3

tion for this divergence can be illuminated by substituting (10) into (15) to obtain:

$$w(z(\tau),\tau)=\left[\frac{E}{\eta(1-\delta)}\right]\left[\frac{1-\delta^{\tau}}{1-(\lambda^{z}\delta)^{\tau}}\right] \tag{16}$$

The change in (16) with respect to τ can be expressed as

$$\frac{dw}{d\tau}=\left(\frac{\partial w}{\partial\lambda^{z}}\right)\left(\frac{\partial\lambda^{z}}{\partial z}\right)\left(\frac{dz}{d\tau}\right)+\left(\frac{\partial w}{\partial\tau}\right). \tag{17}$$

Equation (17) reveals that the change in w with respect to τ is the sum of two distinct components. The first component reflects the indirect influence of τ on w through z. This influence is embodied in the product of three derivatives. The first two elements of this product capture the effect of z on w, which is always positive since z (the rate of innovation) varies positively with consumer utility (see equation 5). The sign of the third element ($dz/d\tau$) depends on the trade-off between size and frequency discussed in the previous section. The second component of (17), $\partial w/\partial\tau$, is the direct effect on welfare of less frequent innovation. This component is always negative due to consumer discounting. Hence, the trade-off between the size and frequency of innovation again determines the maximizing patent length. However, from the welfare perspective, the balance point between size and frequency (τ^{ω}) occurs before τ^*. To see this, simply note that at $\tau=\tau^*$, $dz/d\tau=0$ and welfare are already falling in patent length (since $\partial w/\partial\tau<0$). The divergence between τ^* and τ^{ω} is therefore attributable to consumer discounting. While size affects w only through z, frequency operates through two channels: through z and through consumer discounting.

From a policy perspective, divergence of the patent length that maximizes welfare and that which maximizes the rate of innovation creates a tension in the choice of patent length. If, consistent with the explicit Congressional objective, patent length is chosen to maximize the rate of technological advance, a price in consumer welfare must be paid.

3. CONCLUSION

The objective of this paper is to identify the patent length that maximizes the rate of innovation and compare it with the patent length that maximizes consumer welfare. Our model suggests these patent lengths are distinct, and that the patent length that maximizes the rate of innovation exceeds that which maximizes welfare. Though our model is simple, the fundamental mechanism responsible for the existence of maximizing patent lengths, and their divergence, is intuitive, and should be operative in more complex formulations. Specifically, increases in patent length have countervailing effects on the size and frequency of innovation. A longer patent length increases the size, but decreases the frequency, of innovation. The patent length that maximizes the rate of innovation balances the size and frequency effects. Similarly, the patent length that maximizes consumer welfare balances the size and frequency effects from a welfare perspective. The rate of innovation and welfare balance points differ because of the more powerful effect of frequency on consumer welfare. A decrease in the frequency of innovation reduces welfare directly through discounting, and indirectly through its effect on the rate of innovation. Hence, reduced frequency dominates increased size at a lesser patent length from the welfare perspective.

The divergence in the patent lengths that maximize welfare and the rate of innovation raise a number of important policy questions. In addition to patents, the government employs a complex set of tax and regulatory incentives to influence firms' decisions to innovate. Could these instruments be used to offset the "distortion" created by choosing a patent life that maximizes the rate of innovation rather than consumer welfare? Though analysis of this question is beyond the scope of the current paper, our model provides a framework for broaching such coordination issues. The model can also be employed to explore variations in the maximizing patent length as tastes, production costs, or time preferences change. Such an exercise would be suggestive of how patent lengths might vary across industries in an ideal world. In-depth analysis of these issues will be the object of future research.

Vanderbilt University, U.S.A.
Vanderbilt University, U.S.A.

APPENDIX

PROPOSITION 1. *The extension of z (equation 10) to the real line (call it Z) has the following properties:*

(i) *Z has a unique positive stationary point that is a maximum.*
(ii) $Z \to 0$ *as* $\tau \to \infty$.
(iii) $dZ/d\tau > 0$ *when* τ *is sufficiently small.*

PROOF. (i). Differentiating Z with respect to τ, yields:

$$\text{(A1)} \qquad \frac{dZ}{d\tau} = \frac{1}{\tau \ln \lambda}\left(\frac{-1}{\tau} \ln\left[\frac{(1-\delta^{\tau})E \ln \lambda}{C(1-\delta)}\right] + \frac{\delta^{\tau} \ln \frac{1}{\delta}}{1-\delta^{\tau}} \right)$$

Setting (A1) equal to zero and manipulating yields

$$\text{(A2)} \qquad (\delta^{-\tau}-1)\ln\left[\frac{(1-\delta^{\tau})E \ln \lambda}{C(1-\delta)}\right] = \tau \ln\frac{1}{\delta}$$

Define the LHS of (A2) as the function $LA2(\tau)$, and RHS of (A2) as the function $RA2(\tau)$. To demonstrate the existence of a unique *positive* τ that satisfies (A2), ($\tau = 0$ is a trivial solution to A2), differentiate $LA2(\tau)$ to obtain:

$$\text{(A3)} \qquad \frac{dLA2}{d\tau} = \left(\delta^{-\tau} \ln\left[\frac{(1-\delta^{\tau})E \ln \lambda}{C(1-\delta)}\right] + 1\right) \ln\frac{1}{\delta}$$

By (11), $(dLA2/d\tau) > 0$, for all $Z > 0$. Furthermore $d^2LA2/d\tau^2 > 0$ for all $\tau > 0$, establishing the convexity of $LA2(\tau)$. $RA2(\tau)$ is simply a ray from the origin with slope $\ln(1/\delta) > 0$. Figure A1 plots the $LA2(\tau)$ and $RA2(\tau)$.

Hence, there exists a unique, positive τ^* at which $dZ/d\tau = 0$. From (10) and (A2) we see that

$$\text{(A4)} \qquad Z^* = \frac{\delta^{\tau^*} \ln\frac{1}{\delta}}{(1-\delta^{\tau^*})\ln \lambda}$$

Therefore, Z^* is positive and unique. We now demonstrate that the stationary point, τ^*, is a maximum of Z.

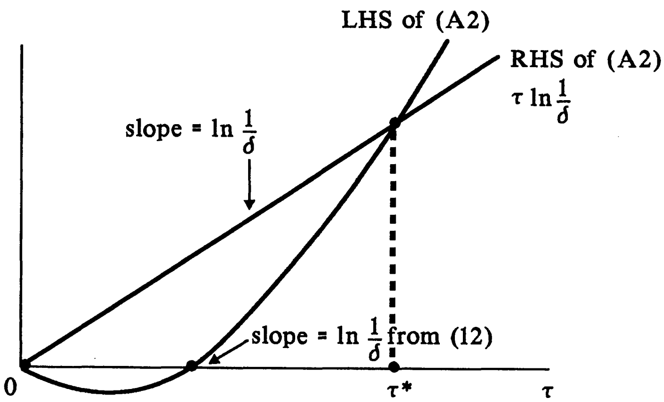

FIGURE A1

(ii) From (10), as $\tau \to \infty$, $Z \to 0$.

(iii) In (A1) it can be seen that as τ is sufficiently small, the term in the square brackets becomes one, and $dZ/d\tau > 0$.

Thus, since τ^* is unique, it must be a maximum and $Z(\tau)$ takes the form illustrated in Figure 1. Q.E.D.

PROPOSITION 2. *The extension of w (equation 15) to the real line (call it W) has the following properties:*

(i) *W has a unique maximum.*

(ii) *$W(0) = 0$ and $W(\tau) = E/\eta(1-\delta)$ as $\tau \to \infty$.*

PROOF. (i) From equation (14) we have

$$W(\tau) = \left[\frac{1-\delta^\tau}{1-\delta}\right]\left(\frac{E}{\eta}\right)\sum_{j=0}^{\infty}\delta^{j\tau}\left[\frac{(1-\delta^\tau)E\ln\lambda}{C(1-\delta)}\right]^j \tag{A5}$$

$$= \left[\frac{1-\delta^\tau}{1-\delta}\right]\left(\frac{E}{\eta}\right)\sum_{j=0}^{\infty}\left[\frac{(1-\delta^\tau)\delta^\tau E\ln\lambda}{C(1-\delta)}\right]^j.$$

Inspection of (A5) reveals that a necessary condition for the existence of a finite W and, hence, a τ that maximizes welfare is

$$(1-\delta^\tau)\delta^\tau E\ln\lambda < C(1-\delta). \tag{A6}$$

Given that (A6) is satisfied, (A5) becomes

$$W(\tau) = \left[\frac{1-\delta^\tau}{1-\delta}\right]\left(\frac{E}{\eta}\right)\frac{1}{1-\left[\dfrac{(1-\delta^\tau)\delta^\tau E\ln\lambda}{C(1-\delta)}\right]}, \tag{A7}$$

which can be rewritten as

$$W(\tau) = \left[\frac{E}{\eta(1-\delta)}\right]\frac{1}{(1-\delta^\tau)^{-1}-\left[\dfrac{\delta^\tau E\ln\lambda}{C(1-\delta)}\right]}. \tag{A8}$$

Now define $\theta(\tau) = (1-\delta^\tau)^{-1} - A\delta^\tau$, where $A = \left[\dfrac{E\ln\lambda}{C(1-\delta)}\right]$. By inequality (A6), $A > 1$. Then we can rewrite (A7) as

$$W(\tau) = \left[\frac{E}{\eta(1-\delta)}\right]\frac{1}{\theta(\tau)}. \tag{A9}$$

The behavior of $W(\tau)$ thus depends on the properties of $\theta(\tau)$. Differentiating θ, we obtain

$$\text{(A10)} \qquad \frac{\partial\theta}{\partial\tau} = -(1-\delta^{\tau})^{-2}\delta^{\tau}\ln\frac{1}{\delta} + A\delta^{\tau}\ln\frac{1}{\delta}$$

which has an extreme point (call it τ^{ω}) where $-(1-\delta^{\tau})^{-2} + A = 0$. Solving for τ^{ω} yields

$$\text{(A11)} \qquad \tau^{\omega} = \frac{1}{\ln\delta}\ln\left[1 - \left[\frac{C(1-\delta)}{E\ln\lambda}\right]^{1/2}\right].$$

Differentiating (A10) and simplifying, yields

$$\text{(A12)} \qquad \frac{\partial^2\theta}{\partial\tau^2} = \delta^{\tau}(\ln\delta)^2\left[(1-\delta^{\tau})^{-2}\left[2(\delta^{-\tau}-1)^{-1}-1\right] + A\right].$$

Evaluating (A12) at τ^{ω} and simplifying, we find

$$\text{(A13)} \qquad \left.\frac{\partial^2\theta}{\partial\tau^2}\right|_{\tau=\tau^{\omega}} = 2\delta^{2\tau}(\ln\delta)^2(1-\delta^{\tau})^{-3} > 0.$$

Equation (A12) can also be used to demonstrate that a unique inflexion point (call it τ') of $\theta(\tau)$ exists:[14]

$$\text{(A14)} \qquad \tau' = -\frac{1}{\ln\delta}\ln\left[1 + \left[\frac{E\ln\lambda}{C(1-\delta)} - 1\right]^{-1/2}\right].$$

With some manipulation it can be shown that $\tau' > \tau^{\omega}$. The final step in our characterization of the function $\theta(\tau)$ function is the evaluation of (A10) at the limits of τ. Using our definition of θ and (A10), we find that

$$\text{as} \quad \tau \to 0, \quad \theta \to \infty, \quad \partial\theta/\partial\tau \to -\infty$$

and

$$\text{as} \quad \tau \to \infty, \quad \theta \to 1, \quad \partial\theta/\partial\tau \to 0.$$

Putting together these properties, the function $\theta(\tau)$ must have the form illustrated in Figure A2 below.

(ii) Now recalling (A9), $W(\tau) \to 0$ as $\tau \to 0$, and $W(\tau) = E/[\eta(1-\delta)]$ as $\tau \to \infty$. Q.E.D.

Hence $W(\tau)$ has the approximate form illustrated in Figure 2 of the text.

[14] A formal proof of the uniqueness of the inflexion point τ' is available from the authors upon request.

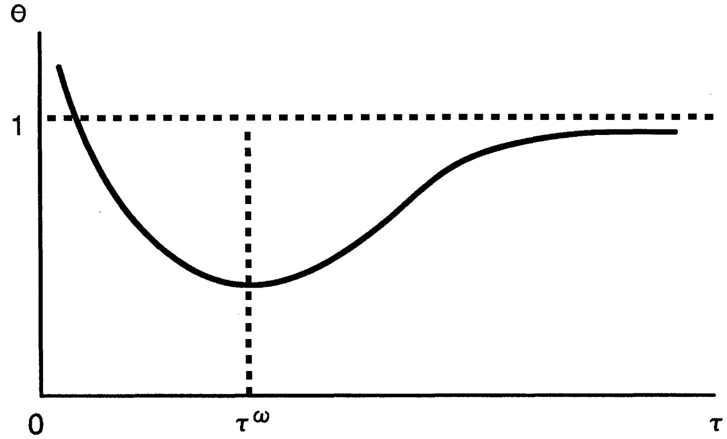

FIGURE A2

PROPOSITION 3. *$\tau^* > \tau^{\omega}$ for all values of B for which τ^* and τ^{ω} are defined (i.e., values of B consistent with inequalities* 11 *and A*6).

PROOF. From (A11), τ^{ω} can be expressed as follows:

$$\text{(A15)} \qquad \tau^{\omega} = \frac{\ln\left[\dfrac{1}{1-B}\right]}{\ln\dfrac{1}{\delta}} \quad \text{where } B = \left[\frac{C(1-\delta)}{E\ln\lambda}\right]^{1/2}.$$

Figure A1 reveals that for any $\tau < \tau^*$, $RA2(\tau) > LA2(\tau)$. Similarly, for any $\tau > \tau^*$, $LA2(\tau) > RA2(\tau)$. Therefore, the direction of inequality between $RA2(\tau^{\omega})$ and $LA2(\tau^{\omega})$ establishes the relationship between τ^* and τ^{ω}. Inequality (11) establishes $B = 1$ as an upper bound on B for which τ^* is defined. Inequality (A6) establishes a lower bound of $B = 1/2$ for which W is finite and τ^{ω} is defined (since $[(1-\delta^{\tau})\delta^{\tau}]^{1/2}$ has a maximum value of 1/2). Utilizing equation (A15) and noting that $\delta^{\tau^{\omega}} = 1 - B$, $LA2(\tau)$ can be expressed as $[-B/(1-B)]\ln B$ and $RA2(\tau)$ as $-\ln[1-B]$. The following inequality thus characterizes the relationship between τ^{ω} and τ^* for $B \in (\frac{1}{2}, 1)$.

$$\text{(A16)} \qquad B\ln B > (1-B)\ln(1-B) \Leftrightarrow \tau^{\omega} < \tau^*$$

Now define $\psi(B) = B\ln B - (1-B)\ln(1-B)$. Note that this function is continuous on the closed interval $B \in [\frac{1}{2}, 1]$. Using L'Hopital's rule $\psi(B) = 0$ at $B = 1$, and by inspection $\psi(B) = 0$ at $B = 1/2$. Differentiating $\psi(B)$ reveals an extreme point at $B = 0.839$. Now note that $\psi''(B) = (1-2B)/[B(1-B)]$, which is negative for all $B \in (\frac{1}{2}, 1)$. Thus $\psi(B)$ is concave and (A16) is always satisfied on the interval $B \in (\frac{1}{2}, 1)$. By (A15), τ^{ω} is convex in B, yielding Figure 3 of the text. Q.E.D.

REFERENCES

AGHION, P. AND HOWITT, "A Model of Growth Through Creative Destruction," *Econometrica* 60 (1992), 323–351.

CHAING, A. C., *Fundamental Methods of Mathematical Economics*, 3rd ed (New York: McGraw-Hill, 1984).

CHOU, C.-F. AND O. SHY, "The Crowding-out Effects of Long Duration of Patents," *RAND Journal of Economics* 24 (1993), 304–312.

GALLINI, N. T., "Patent Policy and Costly Imitation," *RAND Journal of Economics* 23 (1992), 52–63.

GILBERT, R. AND SHAPIRO, C., "Optimal Patent Length and Breadth," *RAND Journal of Economics* 21 (1990), 106–112.

GROSSMAN, G. AND E. HELPMAN, *Innovation and Growth in the Global Economy* (Cambridge: MIT Press, 1992, Chapter 4).

JOVANOVIC, B. AND R. ROB, "Long Waves and Short Waves: Growth Through Intensive and Extensive Search," *Econometrica* 58 (1990), 1391–1409.

JUDD, K., "On the Performance of Patents," *Econometrica* 53 (1985), 567–585.

KLEMPERER, P., "How Broad should the Scope of Patent Protection Be?" *RAND Journal of Economics* 21 (1990), 113–130.

MOOKHERJEE, D. AND D. RAY, "On the Competitive Pressure Created By the Diffusion of Innovations," *Journal of Economic Theory* 54 (1991), 124–147.

NORDHAUS, W., *Invention, Growth and Welfare*, (Cambridge: MIT Press, 1969).

SAATY, T., *Optimization in Integers and Related Extremal Problems* (New York: McGraw-Hill, 1970).

SCHUMPETER, J. A., *Business Cycles* (New York: McGraw-Hill, 1939).

SEGERSTROM, P. S., "Innovation, Imitation and Economic Growth," *Journal of Political Economy* 99 (1991), 807–827.