A Minimum Optimal Patent Term

John F. Duffy*

Most investigations of the optimal patent term problem have followed Nordhaus's assumption that setting an optimal patent term requires balancing the incentives necessary to encourage innovation against the inefficiencies associated with longer lasting monopoly rights. Nordhaus, however, relied on a static model in which all investments and innovations occur at a fixed time. If the times of investment and innovation are not fixed, the time of patent expiration becomes a "U-shaped" function of patent life. Below a certain minimum patent term, increasing patent life results in both earlier innovation and earlier patent expiration. Increasing patent life up to this term involves no tradeoff but instead unambiguously increases welfare.

1. Introduction

The patent system exists to increase social welfare by fostering investment in valuable innovations. The central policy issue in structuring a patent system is to determine the optimal patent right, and one important variable to be optimized is the duration of the patent.

The problem of fixing an optimal patent term has been widely studied and is generally viewed as requiring a balance between the incentives necessary to encourage innovation and the inefficiencies associated with the exclusive right. Nordhaus (1969, p. 76) provides the classic introduction to the problem:

> What are the considerations that determine the optimal life of a patent? As the life is increased, two opposite forces affect the level of economic welfare. First, a longer life increases invention and thus gives on balance a larger amount of output for a given level of inputs. This is a positive effect. Second, a longer life means that the monopoly on information lasts longer and thus there are more losses from

inefficiencies associated with monopoly. The optimal life of a patent is that point at which the two forces balance at the margin.

Other researchers have made similar assumptions, and the general consensus is that increases in patent term "defer" or "postpone" the time at which monopoly distortions are eliminated and society captures the full benefits of innovation.[1] This view of the optimal patent term problem is, however, based on the Nordhaus model which takes a "static" approach to innovation (see Nordhaus at 18): It assumes (i) that all investments in research occur at a fixed time $t = 0$, and (ii) that those investments bear fruit immediately so that the time of invention is also fixed at $t = 0$. The model permits a longer patent term to increase the "size" of innovation but not to affect the *timing* of innovation. Thus, increasing patent life in the Nordhaus model unambiguously pushes the date of patent expiration later and thereby extends monopoly power into later periods. The relationship between patent term and patent expiration is therefore strictly linear, as shown in Figure 1.

In contrast to the Nordhaus model, models such as those of Barzel (1968), Loury (1979) and Dasgupta and Stiglitz (1980b) do not assume the time of innovation to be fixed but instead treat it as a function of a number of variables including the term of patent protection. Where the timing of innovation is allowed to vary, the time of patent expiration will, under a broad range of assumptions, be a "U-shaped" function of patent term, as shown in Figure 2. In such dynamic models, increases in patent term can lead to an acceleration, not to a deferral or postponement, of society's capture of the full benefits of the innovation.

For large values of L ($L > L^*_{min}$) in Figure 2, the basic tradeoff identified by Nordhaus does occur: The increase in patent life produces the innovation slightly sooner but also postpones the expiration of the monopoly right. Indeed, at the very top portion of Figure 2, the decrease in innovation time is so slight that the time of patent expiration approaches a linear function of patent term, just as in Figure 1. Yet an increase in patent term does not necessarily involve this tradeoff. For small values of L ($L < L^*_{min}$), increasing the patent term causes both the time of innovation *and* the time of patent expiration to move earlier. In this region, increasing patent life creates no tradeoff. The earlier expiration of the patent unambiguously increases welfare, and the earlier innovation time may also increase welfare if consumers realize a surplus during the patent term.

The case shown in Figure 2 — i.e., a case where the time of patent expiration (t_E) moves earlier for some increases to patent life — will occur in any model of innovation provided that:

(1) the time of innovation (t_I) goes to infinity as patent term[2] goes to zero ($t_I(L) \to \infty$ as $L \to 0$);

(2) the time of innovation is finite for positive values of patent length ($t_I(L) \neq \infty$ for $L > 0$); and

(3) the time of innovation is a continuous function of L for $L > 0$.

Under these assumptions, $\partial t_I(L)/\partial L \to -\infty$ as $L \to 0$, and thus the function $t_E(L) = t_I(L) - L$ will have a negative slope (negative $\partial t_E/\partial L$) for some positive values of L. If we assume further that $\partial t_E(L)/\partial L$ is positive for some positive values of L,[3] then there will be some patent term L^*_{min} that will produce the earliest date of patent expiration.

The patent term L^*_{min} is worthy of study for a variety of reasons. First, increasing patent term up to L^*_{min} involves no tradeoff of the sort described by Nordhaus because the longer patent term moves the time of patent expiration earlier.[4] This observation does not contradict what has been termed the "basic dilemma of the patent system" — that the patent system creates monopolistic distortions in order to encourage innovation. Tirole (1985), at 390. The tradeoff between monopoly distortion and greater innovation remains central to the policy question of whether to have a patent system rather than, for example, a system of social subsidies, which would not have monopolistic distortions (but would have other drawbacks). However, once society has chosen to rely on a patent system for innovation, social welfare increases with all increases of patent term up to L^*_{min}. A corollary is that the patent length L^*_{min} constitutes a minimum optimal patent length: The optimal patent term for any given invention will be equal to, or longer than, L^*_{min} without regard to the deadweight loss caused by the patent. Another corollary is that, if the patent term is set at or below L^*_{min}, innovations cannot occur inefficiently early and thus the "over-racing" problem identified by Barzel (1968) cannot arise.[5]

Second, this lower bound for an efficient patent term may be much longer than previously thought. Relying on the Nordhaus model, others have shown that if the competition to obtain patents is allowed to dissipate 100% of patent rents, then the optimal patent term is very short — i.e., "less than six months." Berkowitz and Kotowitz (1979), at 10; see also McFetridge and Rafiquzzaman (1986), at 93. But these short optimal patent terms are directly attributable to the static assumptions in the Nordhaus model. With dynamic assumptions, the minimum optimal patent term can be shown (using standard assumptions drawn from the existing literature) to be between one to three decades, with the optimal term longer still if the invention has "spillover" benefits (i.e., benefits captured by persons other than the patentee) during the patent term.

Third, for inventions having no "spillover" benefits during the life of the patent, the problem of an optimal patent term can be recast as simply one of minimizing the time of patent expiration, and L^*_{min} is then the optimal patent term. The size of the patent's deadweight loss is irrelevant, and the tradeoff identified by Nordhaus never occurs. Since the literature on innovation suggests that inventions of this class are quite common (Nordhaus and others describe them as "run-of-the-mill" inventions), L^*_{min} may be the optimal patent term for a broad class of inventions.

Fourth, prior investigations of the patent term problem have relied on either deterministic or stochastic models of innovation. This paper uses both models and shows that both produce similar results concerning the optimal patent term, provided that the two models make the same assumptions about *time*. Specifically, if both models assume that the times of investment and of innovation may vary as a function of patent length, then the models produce similar results. In prior analyses, the difference between stochastic and deterministic models has seemed exaggerated because the deterministic models were invariably static while stochastic models were at least partially dynamic.

Fifth, the models facilitate investigation of the effects of deviating from the optimal patent term. Under the models, the efficiency of patent system is relatively insensitive to modest (10-20%) deviations from the optimal term but is substantially affected by larger deviations, particularly by deviations toward an inefficiently short term. These results suggest that patent terms as short as six months or a year are unlikely to be observed in the real world. The results contrast sharply with those of Nordhaus's model, which predicts that terms of one-quarter to one-seventh of the optimal can still generate 90% or more of all possible welfare gains from a patent system.

Sixth, this analysis also suggests new avenues for research. For example, though prior work has considered the problem of setting an optimal patent term, no work has considered the question of the optimal rate of change of a patent term — i.e., the optimal rate by which a government should increase patent terms if the government determines that patent terms are suboptimal. This question may be especially important where a country is first instituting a patent system, or where the patent system is being extended into a new field (such as business methods).

Below two models are set forth. The first uses a deterministic model of innovation; the second, a stochastic one. Both models assume that innovation takes place in a growing economy;

both are fully dynamic in that they allow the times of invention and innovation to vary; and both yield the same explicit formula for L^*_{min}. The other assumptions made in the models are admittedly simple but are drawn from the existing literature. The consequences of relaxing the assumptions will be investigated in section 6 below.

2. A Deterministic Model

The first model of innovation follows Nordhaus in assuming that a certain research investment *I* will instantaneously yield, with probability 1, an innovation with positive social value. (The assumption of instantaneous invention is not essential. The variable *I* can also be considered the present value, at time t, of a flow of research expenditures necessary to produce the invention at time t. See Dasgupta and Stiglitz (1980b), at 6 n.7.) Also as in Nordhaus, the innovation is assumed to be protected by a patent of duration *L* beginning at the time of innovation, and all pecuniary flows are subject to a constant discount rate *r*. The model also assumes, as Nordhaus does, that the rewards afforded by the patent system will induce efforts directed toward innovation.

The model departs from Nordhaus in three significant respects. *First*, the model assumes that the market for innovation is competitive and thus, in equilibrium, the profits of innovating firms are zero ($\pi_i = 0$). By contrast, Nordhaus assumed that an innovating firm would be able to choose its level of research investment so as to maximize profits (i.e., the firm could choose I such that $\partial\pi_i/\partial I = 0$). As others have pointed out (Stoneman (1987) and others), Nordhaus's assumption on this point does not permit firms to compete in trying to capture the profits from innovation. A competitive market for innovation seems a more realistic assumption. While the implications of competition are not entirely clear, this paper follows Dasgupta and Stiglitz (1980b) in assuming that competition will drive profits to zero and that a single firm will make the investment in innovation.[6]

Second, Nordhaus took the "size" of an innovation to be endogenous to the system; as more resources were devoted to innovation, the size of the innovation increased. In this paper, the size of an innovation will be taken as fixed — i.e., innovation is taken as *quantized*. As more resources are devoted to innovation, the size of each innovation quantum remains unchanged but occurs earlier in time. This assumption is consistent with the approach taken in Loury, Dasgupta and Stiglitz and Denicolo. It is also consistent with the law of patents, which confers a patent right where inventors have reached a certain legally defined quantum of technical achievement (variously called an "inventive step" or a "nonobvious" advance over the prior art). Where parties are racing in time to obtain patents, it is not unrealistic to assume that they will apply for a patent as soon as they achieve the legally set quantum of technical accomplishment.

Third, Nordhaus relied on a static model in which all invention, and all investment in invention, occurred at a fixed time, t = 0. If an economic model is to account for the observed reality that investments in different innovations occur at different times, then the costs and the benefits associated with innovation cannot both remain constant in time. The model in this paper follows prior literature (see, e.g., Barzel (1968), Kitti (1973), Weeds (2002)) in assuming that the costs of innovation remain constant but that the benefits grow at a rate g, which can be thought of as the rate of economic growth throughout the whole economy and is assumed to be less than the discount rate r. Thus, can we define the maximum flow of social benefits from the invention (i.e., the flow of social benefits that would result if the innovation exists and is not subject to any patent right) as:

$$SB_{\max}(t) = S_o\, e^{g t}$$

where S_o is an arbitrary constant and t is any positive or negative time (time t = 0 being arbitrary, see discussion part 6.a, infra). The intuition here is that the social value of an innovation increases as demand grows. Consider, for example, an innovation that decreases the per unit cost of a process by a. Where production is at level w units, the innovation has social value $a \cdot w$. If production rises to $w + \epsilon$, the social value rises to $a \cdot (w + \epsilon)$. The assumption of growing benefits from innovation thus follows from a "public good" aspect of intellectual property — it can be used any number of times at zero additional social cost.[7]

While assuming constant costs of innovation is a simplification, it is less troubling than it might first appear. The cost of creating any particular innovation might, at first blush, seem to be declining with time, sometimes dramatically so. Thus, for example, the cost of producing a 1 gigahertz computer chip might seem to be dramatically higher in 1975 than in 1995 (when technology was just below this level). But this is not the correct comparison. Achieving a gigahertz chip requires achieving a number of innovation quanta or steps, and in 1975 many more steps remained to be accomplished than in 1995. Under this "ladder" view of innovation, it is not wholly unrealistic to assume that the social cost for achieving each quantum or step in the ladder remains constant. The falling cost of achieving any particular technological level would then be properly attributed to society's progress on the ladder, not to a decrease in cost for each innovation quantum.

If we assume that SB_{max} is the flow of social benefits if the innovation were freely available, these social benefits can be divided into three components during the patent term:

$$SB_{max}(t) = H(t) + J(t) + K(t)$$

where $H(t)$ is the patentee's flow of rents, $J(t)$ is the "spillover" realized by others (e.g., consumers and competitors), and $K(t)$ is a deadweight loss not realized while the patent remains in force. For simplicity, H, J and K are also assumed to increase at rate g, and so the potential social benefit is:

$$SB_{max}(t) = H_o\, e^{gt} + J_o\, e^{gt} + K_o\, e^{gt}$$

H_o, J_o and K_o are nonnegative constants specific to each invention. The patentee's flow of rents will be less the maximum flow of social benefits ($H < SB_{max}$) except where J_o and K_o are zero (e.g., in the theoretical case of perfect price discrimination).[8]

If the cost of innovating (I) is constant in time, the innovator's discounted profits will be:

$$\pi_I(t_I) = \int_{t_I}^{t_I+L} H_o e^{-(r-g)t} dt - Ie^{-rt_I}$$
$$= \frac{H_o e^{-(r-g)t_I}}{(r-g)}\left[1 - e^{-(r-g)L}\right] - Ie^{-rt_I}$$

where t_I is the time of investment and (by assumption) also of innovation. Competition to innovate is assumed to drive profits to zero, and this condition defines the time of investment:

(1) $$0 = \frac{H_o e^{-(r-g)t_I}}{(r-g)}\left[1 - e^{-(r-g)L}\right] - Ie^{-rt_I}$$

(2) $$t_I = \frac{1}{g}\ln\left[I(r-g) / H_o(1 - e^{-(r-g)L})\right]$$

As can be seen from (2), $t_I \to \infty$ as $L \to 0$, and $\partial t_I / \partial L < 0$ for all positive values of L. Thus, as patent term is increased, the time of investment/innovation always decreases.

The innovation's contribution to social welfare, W, will be as follows:

(3) $$W(L) = \int_{t_I}^{t_I+L} (H_o + J_o)e^{-(r-g)t} dt + \int_{t_I+L}^{\infty} (H_o + J_o + K_o)e^{-(r-g)t} dt - Ie^{-rt_I}$$

The zero profit condition (1) means that the total flow of rents to the patentee during the patent term will just equal the investment necessary to create the invention (i.e., all producer surplus is dissipated in the race to achieve the patent), and so equation (3) can be simplified to:

(4) $$W(L) = \int_{t_I}^{\infty} J_o e^{-(r-g)t} dt + \int_{t_I+L}^{\infty} (H_o + K_o)e^{-(r-g)t} dt$$

(5) $$W(L) = \frac{J_o e^{-(r-g)t_I}}{r-g} + \frac{(H_o + K_o)e^{-(r-g)(t_I+L)}}{r-g}$$

Equation (5) is easily understood in terms of the welfare effects of the patented innovation. The first term on the right hand side represents the spillover effects that society captures during the patent term. This increased flow of benefits begins at the time of innovation and continues for all time. Thus, this term always increases as the time of innovation occurs earlier. Since $\partial t_I / \partial L < 0$ for all positive values of L, this first term always increases with increasing L.

The second term in the formula represents the flow of benefits that society realizes only after the patent expires. This term incorporates both the flow of rents previously realized by the patentee (*H*) and the elimination of the deadweight loss (*K*). Equation (5) can be rewritten as:

(6) $$W(L) = \frac{J_o e^{-(r-g)t_I}}{r-g} + \frac{(H_o + K_o)e^{-(r-g)t_E}}{r-g}$$

where $t_E = t_I + L$ and both t_E and t_I are functions of L. The first term always increases with increasing L; the second term increases with L if and only if changes in L move the time of patent expiration earlier — i.e., if and only if $\partial t_E / \partial L < 0$. Thus, social welfare always increases with increasing patent term if the time of the patent expiration is also moving earlier:

(7) $$\partial W / \partial L > 0 \quad \text{if} \quad \partial t_E / \partial L < 0 .$$

The patent term at which $\partial t_E / \partial L = 0$ thus establishes a lower bound for the optimal patent term. Indeed, where there are zero spillovers during the patent term (J = 0), the optimal patent term is the term that minimizes the patent expiration date (i.e., the term for which $\partial t_E (L) / \partial L = 0$).[9] This minimal patent term (L^*_{min}) can be calculated as follows:

(8) $$\partial t_E (L^*_{min}) / \partial L = 0 \quad \text{or, equivalently,} \quad \partial t_I (L^*_{min}) / \partial L + 1 = 0$$

Using equation (2) for the time of innovation, we calculate $\partial t_I / \partial L$ and then have:

(9) $$1 - \frac{r-g}{g}\left(\frac{e^{-(r-g)L}}{1-e^{-(r-g)L}}\right) = 0$$

(10) $$L^*_{\min} = \frac{1}{r-g}\ln\left(\frac{r}{g}\right)$$

Table 1 provides $L^*_{\min}$ for various values of r and g. Interestingly, $L^*_{\min}$ has a fairly good fit with observed patent terms, which have ranged between 14 and 21 years in various nations across the past two centuries. The fit is especially good where g is in the range of 2-5% and the discount rate is moderately high (~10%). A relatively high discount rate may be appropriate since, even though the model here assumes certainty, investments in innovation are in fact quite risky.

As shown in the appendix, a general expression for the optimal patent term, L*, is:[10]

(11) $$L^* = L^*_{\min} + \frac{1}{r-g}\ln\frac{1}{1-\frac{J_o}{H_o+K_o}\frac{r-g}{g}}$$

This expression demonstrates that L* always equals $L^*_{\min}$ provided that $J_o = 0$. This result accords with the intuition set forth above. It also shows that L* goes to ∞ as $J_o \rightarrow g(H_o + K_o)/(r-g)$ — i.e., the optimal patent term become infinite where spillovers allow society to capture immediately (during the patent term) a significant fraction of the overall social benefits of the invention. Again, this result is consistent with our intuition: The limited patent term reduces the patentee's ability to appropriate the social value of the invention; it is designed to curb the excessive patent racing that would otherwise occur. As spillover effects diminish appropriability, the need for further restricting appropriability by limiting patent term also diminishes and eventually vanishes.

3. A Stochastic Model.

Modern models of patent races more commonly consider innovation as a stochastic process. One standard model (see Dasgupta and Stiglitz (1980b); see also Loury (1979)) assumes that an investment in innovation x purchases a constant probabilistic rate of achieving research success, $\lambda(x)$. However, like the Nordhaus model, the Dasgupta and Stiglitz stochastic model assumes that all investment in innovation occurs at a fixed time, t = 0. Thus, the model does not permit firms to compete in the timing of their investments. The Dasgupta and Stiglitz model also assumes that the flow of social return on invention does not vary with time, and so that model is static in a fundamental sense.

This paper departs from the Dasgupta and Stiglitz model in two ways. First, as in the deterministic model, the maximum flow of social benefits from the invention (i.e., the flow that would result if the innovation were freely available) is assumed to increase at rate g:

(12) $$SB_{\max} = S_o\, e^{g t}$$

The intuition here is the same as for the deterministic model. As in the deterministic model, this paper will assume that the cost of innovation (here, $\lambda(x)$) is constant in time.

A second departure from the Dasgupta and Stiglitz model is that firms in this model are permitted to vary the time at which they invest in innovation. In the Dasgupta and Stiglitz model, all investment occurs at time t = 0. In their model, patent racing occurs because more valuable inventions attract *more firms* to make *larger* investments at time t = 0; racing does not occur by firms making investments *earlier*. Thus, the intensity of the patent race in the Dasgupta and Stiglitz model is greater for higher valued inventions than for lower valued inventions, where the invention's value is measured at t = 0. If firms are permitted to vary the time of investment, then a more valuable invention will attract investment sooner than a less valuable invention but, at the time that investment is made, the expected returns of each invention just equal its expected costs.

Allowing firms to race in time also requires some adjustment to the assumptions about the number of firms that will invest. In the Dasgupta and Stiglitz model, a more valuable invention can attract more investing firms as well as larger investments by those firms. Where firms are permitted to race in time, a more valuable invention will attract investment *earlier* but may not increase the number of firms making the investment or the size of the investment made by the firms. For simplicity, this paper will assume that the number of firms capable of investing in the innovation, *n*, is fixed and that the race occurs only in timing of the investment. This assumption may be seen as the reverse of the assumption in the Dasgupta and Stiglitz model, which postulates a fixed time for investment (t = 0) but allows *n* to vary.

With these assumptions, the profits of a firm investing in innovation will be:

$$\pi(t_I)=\int_{-\infty}^{\infty} P(t_I,t,x)V(t)dt-xe^{-rt_I} \tag{13}$$

where t_I is the time of the investment x; $P(t_I, t, x)$ is the probability density function of the first research success occurring at time t where investment x has occurred at time t_I; and V(t) is the value of the patent where success in creating the invention occurs at time t. The probability density function of the first research success then takes the following form:

$$P(t_I,t,x)=\begin{cases} 0 & \text{for } t < t_I \\ \lambda(x)\exp\left[-\sum_{j=1}^{n}\lambda(x_j)(t-t_I)\right] & \text{for } t \geq t_I \end{cases} \tag{14}$$

In other words, the probability of any success is zero prior to the investment. By making an investment of *x*, the firm purchases a constant probabilistic rate of achieving research success, $\lambda(x)$ — i.e., $\lambda\Delta t$ is the probability that a research success will occur during an infinitesimal increment of time Δt. The probability that any particular firm has not yet achieved a research success by time t $\geq t_I$ is then $\exp\{-\lambda(x_j)(t-t_I)\}$. If, as in the Dasgupta and Stiglitz model, we suppose identical firms, we can search for the symmetric Nash equilibrium defined by:

(15)
$$P(t_I,t,x)=\begin{cases} 0 & \text{for } t < t_I \\ \lambda(x)\exp\left[-n\lambda(x)(t-t_I)\right] & \text{for } t \geq t_I \end{cases}$$

Firm profits can be rewritten as:

(16)
$$\pi(t_I)=\int_{t_I}^{\infty}\lambda(x)e^{-n\lambda(x)(t-t_I)}V(t)dt-xe^{-rt_I}$$

The value of the patent is:

(17)
$$V(t) = \int_{t}^{t+L} H_o e^{-(r-g)t}dt = \frac{H_o e^{-(r-g)t}}{(r-g)}\left[1 - e^{-(r-g)L}\right]$$

where the terminology is the same as before — i.e., H_o is the patentee's flow of rents at an arbitrary time zero, g is the rate of growth in rents, and L is the patent term. The innovator's profits are then:

$$\pi(t_I) = \lambda(x)e^{n\lambda(x)t_I}\,\frac{H_o(1 - e^{-(r-g)L})}{(r - g)}\int_{t_I}^{\infty} e^{-[n\lambda(x)+r-g]t}dt - xe^{-rt_I}$$

(18)
$$\pi(t_I) = \left[\frac{\lambda(x)H_o(1 - e^{-(r-g)L})}{(r - g)(n\lambda(x) + r - g)}e^{-gt_I} - x\right]e^{-rt_I}$$

As in the Dasgupta and Stiglitz model, we assume that firms will choose x so as to maximize profits and that competition drives firm profits to zero. These two conditions define, respectively, the level of investment and the time of investment. Importantly, the level of investment does not depend on the duration of the patent or the size of the innovation. We can see this point by finding x, as determined by the profit maximization condition:

(19)
$$\frac{\partial \pi}{\partial x}=0=\left[\frac{\lambda H_o(1-e^{-(r-g)L})}{(r-g)(n\lambda+r-g)}e^{-gt_I}\left(\frac{\lambda'}{\lambda}-\frac{n\lambda'}{n\lambda+r-g}\right)-1\right]e^{-rt_I}$$

(20)
$$0=\frac{\lambda H_o(1-e^{-(r-g)L})}{(r-g)(n\lambda+r-g)}e^{-gt_I}\left(\frac{\lambda'}{\lambda}-\frac{n\lambda'}{n\lambda+r-g}\right)-1$$

The zero profit condition requires that $\pi = 0$, so from equation 18:

(21)
$$\frac{\lambda(x)H_o(1-e^{-(r-g)L})}{(r-g)(n\lambda(x)+r-g)}e^{gt_I}=x$$

Substituting (21) into (20), we have an expression for x that depends only on r, g, n and $\lambda(x)$:

(22)
$$\frac{1}{x}=\frac{\lambda'}{\lambda}-\frac{n\lambda'}{n\lambda+r-g}$$

Though it may at first seem surprising that the level of investment is independent of the patent reward (i.e., independent of L and H), there is a simple intuition underlying this result: Equation (22) means that each firm will choose the optimally sized research project, where the optimal size is determined by the research project that will yield the greatest expected value of patent rents. This optimally sized project will be determined by the number of firms making similar investments, the discount and growth rates, and technological factors (the slope and elasticity of the research success function). It will not depend on the size of the patent rewards since, due to rivalry, there is always a fixed relationship between the expected patent rewards for an invention and the level of investment — i.e., expected rewards are always equal to investment. Because this relationship holds without regard to (i) the invention's level of appropriability (H) and (ii) the length of patent term (L), the optimal level (not the optimal timing) of investment can be stated without regard to these variables.

From equation (21), we can now solve for the time of investment:

(23) $$t_I = \frac{1}{g}\ln\left[\frac{x(r-g)(n\lambda(x)+r-g)}{\lambda(x)H_o(1-e^{-(r-g)L})}\right]$$

This equation is similar to the expression for innovation/investment time in the deterministic model (see equation 2) except that I is replaced with $x(n\lambda + r - g)/\lambda$, which might be considered the "cost" of inventing in the stochastic setting. Because n, r, g and λ are assumed to be independent of L, and because x is independent of L as shown in equation (22), t_I can be expressed in two terms, one of which is constant in L and one of which is not:

(24) $$t_I = \frac{1}{g}\ln\left[\frac{x(r-g)(n\lambda(x)+r-g)}{\lambda(x)}\right] - \frac{1}{g}\ln\left[H_o(1-e^{-(r-g)L})\right]$$

The time of investment depends on L in precisely the same way it did in the deterministic model, and $\partial t_I/\partial L$ is the same as in the deterministic model. In the stochastic model, however, t_I is only the time of investment not, as in the deterministic model, the time of investment *and innovation*. Thus, the time of patent expiration in the stochastic model is not simply equal to t_I plus L.

To determine the optimal patent term, we must calculate the welfare gain from the patent system as a function of L. The increase in welfare associated with invention can be expressed as:

(25) $$W(L) = \int_{-\infty}^{\infty} nP(t_I,t,x)S(t)dt - nxe^{-rt_I}$$

where P is the probability function defined earlier and S(t) is the social value of innovation where the innovation occurs at time t. Applying the definitions of H, J and K from earlier in the paper, we can express the social value of the innovation as:

(26) $$S(t) = \int_t^{\infty}(H_o + J_o)e^{-(r-g)t}dt + \int_{t+L}^{\infty}K_o e^{-(r-g)t}dt$$

(27) $$S(t) = \frac{(H_o + J_o)e^{-(r-g)t}}{r-g} + \frac{K_o e^{-(r-g)(t+L)}}{r-g}$$

Substituting this into the welfare equation and integrating, we have:

(28) $$W(L) = \frac{n\lambda(x)e^{-(r-g)t_I}}{n\lambda(x)+r-g}\left[\frac{(H_o+J_o)}{r-g}+\frac{K_o e^{-(r-g)L}}{r-g}\right]-nxe^{-rt_I}$$

Using the zero profit condition above, we can rewrite welfare as:

$$W(L) = \frac{n\lambda(x)e^{-(r-g)t_I}}{n\lambda(x)+r-g}\left[\frac{(H_o+J_o)}{r-g}+\frac{K_o e^{-(r-g)L}}{r-g}\right]-\left(\frac{n\lambda(x)e^{-(r-g)t_I}}{n\lambda(x)+r-g}\right)\left(\frac{H_o}{r-g}\right)(1-e^{-(r-g)L})$$

(29) $$W(L) = \frac{n\lambda(x)}{n\lambda(x)+r-g}\left[\frac{J_o e^{-(r-g)t_I}}{r-g}+\frac{(K_o+H_o)e^{-(r-g)(t_I+L)}}{r-g}\right]$$

This equation is equivalent to equation (6) in the deterministic model. Indeed, in the limit as $\lambda \to \infty$ (i.e., as the stochastic model approaches the deterministic model), equation (29) becomes identical to (6). Moreover, the first factor in equation (29) is constant in L because n, r, g and λ are assumed to be independent of L and equation (22) shows x to be independent of L. Since the bracketed terms in (29) behave just as the terms in the deterministic welfare expression (see equation (6)), the derivative of the welfare expression $\partial W/\partial L$ differs from $\partial W/\partial L$ in the deterministic model by a constant. Thus, the value of L that maximizes W (the value at which $\partial W/\partial L = 0$) is the same in both models, with the same the optimal patent term:

(30) $$L^* = \frac{1}{r-g}\left[\ln\frac{r}{g}+\ln\frac{1}{1-\frac{J_o}{H_o+K_o}\frac{r-g}{g}}\right]$$

Similarly, L^*_{min} is also identical for the deterministic and stochastic models.

The broad equivalence between the deterministic and stochastic models is to be expected. The deterministic model yields solutions that are independent of the total "size" or full social value (H + J + K) of the invention. The change to the stochastic model can be viewed as simply changing the size of the invention because the innovation's private and social values in the stochastic model are equal to the private and social values in the deterministic model multiplied by the factor:

$$\frac{n\lambda(x)}{n\lambda(x)+r-g}. \tag{31}$$

This factor is constant in L and t (recall that n is assumed to be fixed; the function λ is assumed not to change with time; and x was demonstrated to be independent of L and t). Thus, the stochastic problem reduces to the deterministic problem where H, J and K are each reduced by the factor in (31). This decrease in size will change the optimal time for the innovation to occur (it will push the optimal time of innovation later), but it does not alter the optimal patent length calculation.

4. The Minimum Optimal Patent Term and an Optimal Patent Policy.

One obvious question to ask about the minimal optimal patent term is whether an optimal patent policy would set the legal patent term based on the minimum optimal term or, alternatively, whether it would be desirable to set the term equal to the optimal term. There are good reasons to think that setting the legal patent term equal to the minimum optimal term, or something only slightly longer, is a sound policy course.

The first reason to limit patent term to the minimum optimal term is that the fully optimal term depends upon $J_0/(H_0 + K_0)$ — i.e., the ratio of (a) the spillovers associated with the invention to (b) the patentee's rents plus the deadweight loss. This ratio, which we will refer to as the spillover ratio, is likely to vary from invention to invention. Thus, if society wanted to set the legal patent term equal to the optimal patent term, the legal system would need to be capable of determining the spillover ratio for each invention and then assigning an optimal patent term to each invention based on that ratio. The prior literature has generally assumed that administrative costs would makes such individualized determinations "infeasible." Kaufer (1989), at 42.

Alternatively, if society were to set for all inventions a single term longer than the minimum optimal term, the patent term would be too short for some inventions (those with high spillovers) and too long for others (those with no spillovers). Society could not be confident that the longer term was better than the minimum optimal term unless society knew the distribution of spillover ratios for all inventions. Again, the costs of obtaining such information are likely to be daunting.

Yet even if the distribution of spillover effects for all inventions were known, the optimal uniform patent term might be not much longer than the minimum optimal term. Existing economic theory suggests that, if an invention is a "run-of-the-mill" invention (a small advance in technology which continues to compete with pre-existing technology), the spillover effects associated with the invention may be close to zero during the patent term. Nordhaus (1969), at 71-72. Empirical studies also suggest that most innovations are such incremental, "run-of-the-mill" inventions. If this is so, then the distribution of spillover ratios for all patented inventions may be skewed toward zero, and for the inventions with negligible spillovers, the minimum optimal patent term is the optimal term. Moreover, for such inventions, large increases in patent term can decrease welfare substantially. Figure 3 and Table 2 show the sensitivity of the welfare function to deviations from the optimal term for zero spillover inventions. Social welfare changes little for small deviations from the optimal term; under reasonable assumptions, deviations of 10-20% (or a few years if the optimal term is between 10-30 years) decrease welfare 9% or less. Larger deviations, however, produce much more substantial declines. Under reasonable assumptions, setting the patent term equal to double the optimum loses at least 25% of the potential welfare gain. Tripling the optimal term captures less than half the potential welfare gain in a relatively high growth economy (i.e., $r/g \approx 2$), and the loss is more severe in economies with lower growth rates.[11]

While patent terms longer than the minimum optimal term decrease the welfare realized from zero spillover inventions, they can increase welfare for inventions with positive spillovers. However, this effect is modest even for inventions with substantial spillover effects. For example, Table 3 and Figure 4 show that, even for an invention with spillover effects substantial enough to justify an infinite optimal patent term (i.e., $J_o = (H_o + K_o)g/(r-g)$), a patent system with a minimum optional term will capture over 70% of the theoretically possible gains that could have been realized from the invention if the patent system had set an optimal (i.e., infinite) patent term for the invention.

Patent terms longer than the minimum optimal term produce much greater welfare gains only for those inventions having very high spillovers. Table 4 and Figure 5 show that, for inventions having very large spillover effects (i.e., $J_o/(H_o + K_o) \to \infty$), a patent system with a minimum optimal term will capture less than half of the welfare gains that could have been realized from the invention if the patent system had an optimal (again, infinite) patent term. Of course, inventions having high spillover effects are precisely the inventions that the patent system is least capable of fostering. It may thus be best not to base the design of the patent system on these inventions but instead to supplement the patent system with other mechanisms better adapted to producing such inventions. Specifically, if society does have information concerning the size of spillovers associated with a particular invention or class of inventions, a superior strategy might be for society to keep the patent term at the minimum optimal term but to subsidize innovations based on the degree of the spillovers.

5. Comparison with Prior Patent Length Calculations

While broadly consistent with observed patent terms, the minimum optimal patent terms reported here are generally longer than those calculated under the Nordhaus model, and the welfare gain associated with innovation is much more sensitive to deviations from the optimal term. The optimal patents are also independent of total innovation size, unlike the optimal patent terms Denicolo found using a stochastic model. These differences flow directly from the differing assumptions about time and competition.

a. Insensitivity to Patent Length under the Nordhaus Model.

Nordhaus's most surprising result is that the welfare gain associated with having a patent system is highly insensitive to the patent term. The insensitivity to patent length can be seen in Figure 6, which is taken directly from Nordhaus's work. (Figure 6 can be usefully compared to Figure 3, which shows the welfare gains for nonoptimal patent terms under the deterministic model in this paper.) This insensitivity is attributable to Nordhaus's assumptions about competition and time that may not correspond to reality.

The right-hand side of Figure 6 shows that social welfare is highly insensitive to excessively long terms in the Nordhaus model. Indeed, in one typical example, Nordhaus notes that increase in patent length from 8 years (roughly, the optimal term) to 1000 years will produce, in his model, a change in welfare gains of "about .2 percent." Nordhaus (1969) at 84. This extreme insensitivity to excessive patent life is easily traced to the assumption in the Nordhaus model that an innovator does not face competitive pressure. Thus, an increase in patent life produces mainly a transfer of rents from consumers to the patentee and not a competitive dissipation of rents.

Prior commentators have criticized Nordhaus's assumption of no rivalry in innovation as "highly restrictive," Kamien and Schwartz (1974), at 187, and noted that Nordhaus's model includes "the unrealistic combination of perfect competition in the product market and a monopoly in the innovation market," Berkowitz and Kotowitz (1979). When the assumption of no competitive innovation is replaced with an assumption of competitive dissipation of all patent rents, then the social cost of longer patent terms becomes much larger. As demonstrated by Berkowitz and Kotowitz (1979 at p. 10), adding the assumption of competitive dissipation of patent rents to the Nordhaus model makes "the benefits [of a patent system] decline rapidly past the optimum [patent term]" and also makes the optimal patent length very short — perhaps no more than six months.[12]

The Nordhaus model is also surprisingly insensitive to excessively short patent terms. Indeed, as can be seen from the left hand side of Figure 6, very short patent terms of *six months to two years* can produce 80-90% of all possible welfare gains from a patent system. Nordhaus, at 83-84. Moreover, such short terms become the optimal where Nordhaus's model is modified to include competitive innovation. This leads to the obvious question: Why are short patent terms so valuable in both the Nordhaus model (with its assumption of a monopoly right to innovate) and the modified Nordhaus model (with the assumption of competitive innovation)? Or, equivalently, why does the benefit of increasing the patent term decline so quickly in both models?

The answer is found in the static nature of the Nordhaus model, which assumes that *all* investments in innovation occur at time zero. That assumption means that any increase in patent term, while inducing additional increments of innovation at the margin, also extends into later time periods the monopoly on the infra-marginal increments (the more valuable increments of innovation induced by shorter patent terms).[13] The length of the optimal patent term then depends on the "output elasticity" of research, which is the elasticity of the production savings achieved by the innovation with respect to the investments made in research.[14] Nordhaus assumed that the cost reduction associated with the innovation would be a concave function increasing with the *10^{th} root of the investment*, thus yielding a constant output elasticity of research equal to .1. (Nordhaus at 80.) Under this assumption, the additional investments attracted by longer patent terms suffer diminishing returns at a very rapid rate (e.g., the first unit of investment yields the same percentage savings in production costs as the next 1023 units *combined*). Increasing the patent term creates more marginal increments of innovation but at the high cost of extending the output inefficiencies for the more valuable infra-marginal increments induced by even very short patents. As a result, very short patent terms produce most of the gains possible from a patent system.[15]

In a dynamic model, more valuable inventions are produced earlier in time so that all incremental investments in innovations are, at the time of investment, expected to yield the competitive rate of return earned elsewhere in the economy. In other words, *all investments are marginal at the time they are made.* Moreover, because investment time is allowed to vary, longer patent terms do not necessarily extend the output inefficiencies associated with the patent: The period during which the invention is under the patent right can move earlier in time.

The Nordhaus assumption that all innovation occurs at a fixed time may be most appropriate for modeling a patent system that has just been created. The short optimal patent terms predicted by Berkowitz and Kotowitz under the modified Nordhaus model might be appropriate as the initial term for a new patent system. The intuition here is simple: If a patent system has not previously existed, there might be many available research projects likely to produce valuable inventions. Such projects exist because the patent system was not previously in place and therefore, assuming the innovation can be quickly imitated, the private return on investments in those projects would be too low to justify the investment. When a patent system is first instituted (or perhaps when the existing patent system is first extended to a new field), those projects will attract investment even if the patent term is relatively short. Increasing the term would attract investment to less valuable research projects but at the cost of extending the exclusive rights on the infra-marginal inventions. In such a situation, the basic tradeoff identified by Nordhaus does occur. Indeed, if the market were competitive, the social costs would include not only the extension of the output distortions caused by the patents on the infra-marginal inventions, but also the dissipation of excess rents associated with those patents. The optimal initial patent terms would then be shorter than those calculated in this paper. As time advances, however, the effects caused by the initial creation of the patent system would diminish, and the optimal patent term would tend to the values calculated in this paper.

b. The Denicolo Calculations: Innovation Size and Duplicative Effort

As in the Nordhaus model, the innovation model used by Loury (1979) and Dasgupta and Stiglitz (1980b) requires that all investment in innovation occur at time $t = 0$ but, unlike Nordhaus, the Loury/Dasgupta/Stiglitz model allows the time of innovation to vary. Denicolo (1999) has investigated the optimal patent term for the Loury/Dasgupta/Stiglitz model. The conclusions reached in this paper reinforce some of Denicolo's conclusions and provide additional insight into others. Three points are worth particular attention.

1. *In zero spillover cases, the optimal patent term is found by minimizing the expected time of patent expiration.* In both the deterministic and stochastic models presented above, the term that minimizes the date of patent expiration is the optimal patent term in cases where there are zero spillovers during the patent term ($J = 0$). The intuition behind this result is that, where the patentee

captures all of the rents associated with the invention during the patent term and rivalry for the innovation dissipates all those rents, then the social value of the invention is equal to the discounted value of the rents after patent expiration, which will be maximized by minimizing the date of patent expiration. This result should apply generally to any model that (i) allows the date of invention to vary, and (ii) assumes rivalrous dissipation of all patent rents.

The Loury et al. model meets these conditions and, in fact, Denicolo proves this result for the Loury et al. model, although the result is stated in slightly different terms. Denicolo finds that the optimal patent term for the case of J = 0 is found by maximizing a discount factor (see Denicolo (1999) at p. 838, n.10), which is simply the discount factor for finding the present value of a stream of rents beginning at patent expiration and continuing out to infinity. The discount factor will be maximized when the time of patent expiration is minimized.

2. *The patent term that minimizes the time of patent expiration provides a minimal optimal patent term.* Denicolo's results can also be used to show that the patent calculated for the case of J = 0 (i.e., the patent term that minimizes the time of patent expiration) establishes a minimal optimal patent term provided that the invention "size" — i.e., the flow of private rents during the lifetime of the patent (H) — is held constant. (The reason that H must be held constant in the Denicolo model is discussed in point 3 below.) Thus, if H is held constant, the optimal patent term in Denicolo's calculations is in the form:

$$\text{Optimal Term} = \text{Minimal Optimal Term (J=0)} + \mathscr{F}(J, K)$$

where $\mathscr{F}$ is a function of J and K and is always positive. This result is not expressly stated in the Denicolo article, but it can be easily derived from three of Denicolo's results: (i) The optimal patent term does not depend on K where J = 0 (Denicolo, p. 838 n.10); (ii) the optimal patent term is always increasing in J (id. at 837); and (iii) the optimal patent term is decreasing in K where $J \neq 0$ (id.).[16]

The existence of a minimal term not dependent on deadweight loss is important because it shows the limited function of deadweight loss in determining an optimal patent term. Following Nordhaus, Denicolo introduces the optimal patent term problem in this way: "[I]n deciding the lifetime of a patent, society must balance the gains accruing from faster technological progress against the welfare loss associated with the temporary monopoly in the use of the new technology." Denicolo, at p. 827. Under this view, the deadweight loss (K) should be crucial to the calculation of the optimal patent term for it, for as Denicolo later states, "measures the social cost of patent protection." But both the results of Denicolo's calculations and the results of the model presented in this paper demonstrate that the deadweight loss of the patent is not a social cost of increasing the patent term until the term exceeds some minimum.[17] Thus, deadweight loss does not affect the minimal optimal patent term; it is relevant only for limiting the size of the additional patent term required when spillovers are positive during the patent life (J > 0).

3. *The optimal patent term under the Loury et al. model decreases with increasing invention size because the model does not allow time of investment to vary.* Denicolo proves that the optimal patent term under the Loury et al. model decreases both where H increases and where H, J and K all increase in the same proportion. Thus, the optimal patent term increases for "smaller" inventions. This result is somewhat counterintuitive, and it is roughly the opposite of how actual patent systems are structured.[18] Denicolo "suspect[s] that this particular result depends on the specific functional forms that are implied by the exponential distribution of the date of innovation, and may be reversed with different distributions."

In fact, however, the decrease in the optimal term with increasing invention size can be traced to the assumptions in Loury et al. that the investment time is fixed. That assumption means that, as the size of H (the patent rents) increases, competition to capture the patent right increases the number of firms (n) seeking the patent. And as Denicolo notes (at 838-39 & n.11), the optimal term decreases as n increases ($\partial L^*/ \partial n < 0$). In sum, increasing the size of the invention attracts more duplicative efforts and thus creates a need for greater limits on level of private appropriability.

By contrast, where competitive forces are allowed to push the time of invention earlier and patent rents are smaller in earlier time periods, competition can always push the time of investment back to a point where the expected cost of creating the invention just equals the expected value of the patent rents. Thus, while a "larger," more valuable innovation will correspond to an earlier time of investment (holding all other variables constant), it need not create a greater number of firms seeking the patent and need not create a greater need for reducing the level of private appropriability.

As with the Nordhaus model, the Loury et al. model might be useful for modeling a newly created patent system. The intuition here is the same as for the Nordhaus model: When a patent system is first instituted (for simplicity, let us assume that it is instituted unexpectedly), there will be a variety of research projects that have not previously attracted investment. For any given patent length, some of the available projects will be marginal (barely worthwhile even if only one firm invests in it), while some will be infra-marginal and would yield supra-competitive profits if only one firm were to invest in it. If the patent system had already existed, such infra-marginal projects would have already been "mined out." But otherwise, these infra-marginal projects will attract excessive investment. As discussed below, a similar problem may arise where an exogenous shock creates a variety of new research projects that were previously unavailable.

6. Relaxing the Assumptions.

This section explores qualitatively the implications of relaxing some of the important assumptions of the model.

a. Temporal Boundaries and the Optimal Phase-in Rates. As this paper has stressed, assumptions about time are crucial to the optimal patent term problem. This paper has assumed that both the times of investment and innovation are not fixed and can *always* move earlier. Thus, as noted in section 2 above, the variable t is not restricted to positive numbers. Permitting negative time is not itself problematic since the definition of a time zero is arbitrary in all systems of marking time. What is problematic is that the model assumes a patent system *always has been*, and *always will be*, in place; welfare is maximized by determining the optimal patent term for such an ever-present patent system.

In the real world, patent systems have not always existed. At some point in time, they are "turned on" or, what is similar analytically, patent rights are extended to new fields or are increased in strength or duration. This phenomenon continues even in the modern world. For example, the WTO's Trade Related Aspects of Intellectual Property (TRIPs) agreement requires many poorer countries to institute patent systems by specific deadlines. Even in the United States, patent rights were recently extended to a wholly new area (business methods). Modeling the patent system during such transition periods requires assumptions that lie somewhere between Nordhaus's fixed time assumptions and the assumptions made in this paper. Briefly, the assumptions would then be that patent racing could occur and could force the dates of investment and innovation to move earlier in time (as per the assumptions in this paper) — but not always. For some inventions, the time of investment, and hence of innovation, could not move earlier than a temporal boundary — i.e., the "present time," or the time when legal rights become available. For such inventions, the time of invention and investment would be effectively fixed (as in Nordhaus), or at least the investment time would be fixed (as in the Loury/Dasgupta/Stiglitz model).

Perhaps because temporal assumptions have not previously been considered with care, the prior literature has not investigated such transition periods. The model presented here could be modified to investigate such transition issues and, in particular, to investigate the optimal phase-in function for increasing patent rights. Under many plausible sets of assumptions, the socially optimal term would tend toward limits equal to the optimal terms suggested by this paper, with the optimal rate of change diminishing as the existing patent term approaches the limit of the optimal patent term. The precise shape of the optimal phase-in function may be a fruitful area of future research.

b. Declining cost of innovation. If innovation were modeled as having a falling cost, optimal patent terms would be shorter than those calculated here. Indeed, it is easy to see that, in the limit where the cost of inventing falls from infinity to zero at a time X, the optimal patent term is clearly zero. In such a case, the invention will arrive at time X with or without the patent system. Granting the patent creates only deadweight loss with no social gain. Cases in which the cost of innovation is falling less dramatically require less dramatic reductions to the optimal terms calculated here.

The point about rapidly declining cost of innovation leads to a closely related point. A central feature of this model is that potential inventors will be able to race in time under smoothly developing conditions (i.e., costs for innovation that are fixed and benefits that steadily increase at the constant growth rate g less than the discount rate r). Rapidly falling costs of innovation, or rapidly rising benefits, would suggest shorter patent terms. Indeed, patent terms of the sort predicted here (between one to three decades) will perform least well in areas where an exongenous shock or other discontinuity leads to rapidly declining costs or rapidly increasing benefits from innovation. Such an exogenous shock might include an event that suddenly creates a new set of consumer needs (e.g., the sudden emergence of the internet as a tool of commerce in the mid-1990's), or a new surprising scientific advance with many new applications (e.g., the discovery of high temperature superconductors). In such cases, many minor innovations might follow the shock, but it would be best if the term of exclusivity for such innovations were kept quite short. In fact, the U.S. patent system does this, albeit in a very crude fashion. The patent law makes "obvious" inventions unpatentable, which in effect decreases the period of exclusivity down to whatever period can be gained by a first-mover advantage, and courts consider the absence of a "long felt need" to be one factor weighing against a finding of "nonobviousness." Thus, the law tries to limit patentability where exongenous shocks would make the twenty-year term of U.S. law inappropriate.

c. Administrative and Error Costs. As in the prior literature, this paper does not take into account the costs of administering the patent system. In general, the administrative costs of *granting* should be ignored. Those costs are highly relevant to determining whether society should have a patent system, but irrelevant for determining the optimal patent term once a system is instituted. The cost of continuing to enforce the patent is relevant and will decrease the optimal life. The problem is similar to determining the optimal number of goods (years) to tax, where the extension of the tax to an additional good (year) adds administrative costs. See Yitzhaki (1979).

One other category of administrative costs — the cost of erroneous patent grants — is also relevant to calculating the optimal term. If a patent is erroneously granted on an existing technology or an obvious technological development, then the deadweight loss associated with that patent can be viewed as an administrative cost of the patent system. Importantly, this administrative cost increases with patent length. If this cost were included in the social welfare calculation, the optimal patent term would tend to be shorter.

d. Prospect Features of the Patent System. In each of the three instances detailed above, relaxing the assumptions of this paper tends to decrease the optimal term. But there is one factor that tends to mitigate these effects. The models set forth in this paper assume that *all* of the investment needed to commence the stream of patent rents must be made *before* the beginning of the patent term. In reality, this is not the case. Patent systems generally allow the patenting of very embryonic technological "prospects" — e.g., mere paper inventions not yet created, or laboratory experiments still needing much development. See Kitch (1977). Because of this "prospect" structure of patent systems, the investment needed to obtain a patent is often only a small fraction of the total amount of research and development costs needed to bring the invention to market and thereby to commence the flow of patent rents.

If this prospect structure of patenting is taken into account, then the total investment needed to commence the stream of rents (I) must be divided into the investment needed to patent (I_P) and the investment needed to commercialize the invention (I_C). Where the investment needed to commercialize an invention is much larger than the investment needed to win the patent award ($I_C >> I_P$), patent racing will make the welfare gains from the patent system somewhat insensitive to overly long patent terms. In such cases, the patent racing could be analogized to a Demsetzian auction for an exclusive franchise. By competing to patent earlier, firms effectively place "bids" to decrease the rents from the exclusive franchise and, as in Demsetzian auctions, the winning bid will be the one that maximizes the social surplus from the franchise — i.e., the one that will expire soonest. If few resources are expended in actually capturing the patent, then an overly long patent term may create little inefficiency. Firms would patent earlier but then delay commercialization longer, so that the actual beginning and end of market exclusivity for the invention would be identical in any patent system with a patent term at or above the optimal. See Duffy (2004b).

e. Obsolescence. This paper assumes that innovation continues yielding rents forever. One classic way to reconcile this assumption with the observed reality of obsolescence is to postulate that "the 'scrapping' of the design of an object does not necessarily imply the scrapping of the knowledge represented by its creation" and that knowledge is instead always "cumulative." Knight (1944).

Yet if knowledge does become obsolete, it is still easy to show that obsolescence has no effect on the optimal patent term *provided* that the time of obsolescence is fixed and *not* a function of the patent term. However, that assumption is highly restrictive because, to be independent of the patent term length, the time of obsolescence must necessarily be independent of the time of invention. The more general case where the time of obsolescence is a function of invention time (and hence of patent length) is more interesting, but it is also outside the scope of this paper.

e. *Patent Breadth.* Finally, it is worth noting that the introduction of a "patent breadth" policy variable does not undermine the minimum optimal patent term calculations made above. This paper has generally assumed that, during the patent term, the patentee and the rest of society will reap, respectively, fractions H and J of the social benefits accruing from the innovation. A reduction in patent breadth (by, for example, permitting imitation or allowing some other use of the innovation)[19] can be accommodated within the model by increasing spillovers (J) and decreasing patentee rents (H) and deadweight loss (K). That change will increase the optimal term but will not affect the minimum optimal term. It is a separate question, which is not investigated here, whether such a reduction in patent breadth, with a corresponding increase in patent life, is desirable.[20]

7. Conclusion

This paper has focused on formulating a theory for a minimal optimal patent term, which is the term that produces the earliest date of patent expiration. Increasing patent terms up to this minimum unambiguously increases social welfare by eliminating the monopoly distortions of the patent right sooner and may also increase welfare by accelerating the time of innovation. The models in this paper have shown that the minimum optimal patent term appears to have a basic correspondence with observed patent terms, which have historically tended to fall between one to two decades. The models also suggest that, while small deviations from the optimal term (in the order of a few years) may not greatly affect the efficiency of the patent system, larger deviations do. In particular, the models explain why very short patent terms of one to two years, which earlier work has suggested could capture most or all of the efficiency gains from a patent system, would not in fact capture a large fraction of the theoretically possible gains from a patent system.

References:

BARZEL, Y. "Optimal Timing of Innovations." *Review of Economics & Statistics*, Vol. 50 (1968), pp 348-355.

BERKOWITZ, M.K., AND KOTOWITZ, Y. "Optimal Patent Design with Free Entry in the Innovation Market." Working Paper No. 7921, University of Toronto Institute for Policy Analysis (1979).

——. "Patent Policy in an Open Economy." *Canadian Journal of Economics*, Vol 15 (1982), 1-17.

DASGUPTA, P. AND STIGLITZ, J. "Industrial Structure and the Nature of Innovative Activity." *Economic Journal*, Vol. 90 (1980a), pp. 266-293.

——. "Uncertainty, Industrial Structure and the Speed of R&D." Bell Journal of Economics, Vol. 11 (1980b), pp. 1-28.

DEBROCK, L. "Market Structure, Innovation and Optimal Patent Life." *Journal of Law & Economics*, Vol. 28 (1985), at pp. 223–244.

DENICOLO, V. "Patent Races and Optimal Patent Breadth and Length." *Journal of Industrial Economics*, Vol. 44 (1996), pp. 249-265.

——. "The Optimal Life of a Patent When the Timing of Innovation is Stochastic." *International Journal of Industrial Organization*, Vol. 17 (1999), pp. 827-846.

DORE, M. H. I., KUSHNER, J., AND MASSE, I. "The Optimal Length of a Patent with Variable Output Elasticity and Returns to Scale in R&D." *Atlantic Economic Journal*, Vol. 21 (1993), pp. 10 – 26.

DUFFY, J.F. "The Marginal Cost Controversy in Intellectual Property." *University of Chicago Law Review*, Vol. 71 (2004a), pp. 37-56.

——. "Rethinking the Prospect Theory of Patents." *University of Chicago Law Review*, Vol. 71 (2004b), pp. 439-510.

GALLINI, N.T. "Patent Policy and Cost Imitation." *RAND Journal of Economics*, Vol. 23 (1992), pp. 52-63.

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

JANIS, M. D. Second Tier Patent Protection. *Harvard International Law Journal*, Vol. 40 (1999), pp. 151–219.

KAMIEN, M. I., AND SCHWARTZ, N. L. "Patent Life and R&D Rivalry." *American Economic Review*, Vol. 64 (1974), pp. 183–187.

KAUFER, E. *The Economics of the Patent System*, New York: Harwood Academic Publishers (1989).

KITCH, E.W. "The Nature and Function of the Patent System." *Journal of Law & Economics*, Vol. 20 (1977), at 265–290.

KITTI, C. "Patent Policy and the Optimal Timing of Innovations." PhD Thesis, University of Chicago Department of Economics (March 1973).

KLEMPERER, P. "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics*, Vol. 21 (1990), pp. 111–130.

KNIGHT, F. H. "Diminishing Returns from Investment." *Journal of Political Economy*, Vol. 52 (1944), pp. 26–47.

LOURY, G. C. "Market Structure and Innovation." *Quarterly Journal of Economics*, Vol. 93 (1979), pp. 395 – 410.

MCFETRIDGE, D. G., AND RAFIQUZZAMAN, M. "The Scope and Duration of the Patent Right and the Nature of Research Rivalry." *Research in Law & Economics*, Vol. 8 (1986), pp. 91–120.

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

ROMANO, R. E. "Benefits of Competition for Patents." *Information Economics and Policy*, Vol. 4 (1990), pp. 31-43.

SCHERER, F.M. "Nordhaus' Theory of Optimal Patent Life: A Geometric Reinterpretation." *American Economic Review*, Vol. 62 (1972), pp. 422–427.

STONEMAN, P. "The Economic Analysis of Technology Policy." Oxford: Oxford University Press (1987).

TAKALO, T. "On the Optimal Patent Policy." Finnish Economic Papers, Vol. 14 (2001), pp. 33-40.

TIROLE, J. *The Theory of Industrial Organization*, Cambridge, Mass.: MIT Press (1989).

WEEDS, H. "Strategic Delay in a Real Options Model of R&D Competition." *Review of Economic Studies*, Vol. 69 (2002), pp. 729-747.

WRIGHT, B.D. 'The Economics of Invention Incentives: Patents, Prizes, and Research Contract." *American Economic Review*, Vol. 73 (1983), 691-707.

WRIGHT, D.J. "Optimal Patent Breadth and Length with Costly Imitation." *International Journal of Industrial Organization*, Vol. 17 (1997), pp. 419 - 436.

YITZHAKI, S. "A Note on Optimal Taxation and Administrative Costs." *American Economic Review*, Vol. 69 (1979), pp. 475–480.

Appendix 1: A General Expression for L*.

Equation 5 in the text gives an expression for W(L):

(1a) $$W(L) = \frac{J_o e^{-(r-g)t_I}}{r-g} + \frac{(H_o + K_o)e^{-(r-g)(t_I+L)}}{r-g} = \frac{e^{-(r-g)t_I}}{r-g}\left[J_o + (H_o + K_o)e^{-(r-g)L}\right]$$

Substituting the value for t_I found in equation 2 from the text, we have:

(2a) $$W(L) = \frac{1}{r-g}\left(\frac{H_o(1-e^{-(r-g)L})}{I(r-g)}\right)^{(r-g)/g}\left[J_o + e^{-(r-g)L}(H_o + K_o)\right]$$

If a constant C is defined as:

(3a) $$C \equiv \left(\frac{H_o}{I(r-g)}\right)^{(r-g)/g}$$

then W(L) can be rewritten as:

(4a) $$W(L) = C\left(1-e^{-(r-g)L}\right)^{(r-g)/g}\left[J_o + e^{-(r-g)L}(H_o + K_o)\right]/(r-g)$$

Differentiating and rearranging terms gives:

(5a) $$\partial W/\partial L = Ce^{-(r-g)L}\left(1-e^{-(r-g)L}\right)^{(r-2g)/g}\left\{\frac{r-g}{g}J_o + \frac{re^{-(r-g)L}-g}{g}(H_o + K_o)\right\}$$

A first order condition for a maximum of W(L) is dW/dL = 0. For $L \neq 0$ or ∞, W will be maximized by L* as defined by the following:

(6a) $$0 = \frac{r-g}{g}J_o + \frac{re^{-(r-g)L^*}-g}{g}(H_o + K_o)$$

Rearranging terms produces equation (11) in the text.

Appendix 2: The welfare effects of suboptimal patent terms for zero spillover inventions.

Define a function $Y = W(xL^*_{min})/W(L^*_{min})$, which is the welfare gain associated with a patent term of xL^*_{min}, where x is any real positive number and the welfare gain at L^*_{min} is normalized to

1. Substituting the value of L^*_{min} from equation (10) into the general expression for welfare as a function of L found in equation (4a) above, we have:

(7a)
$$Y(x)=\left[\frac{1-\left(\frac{g}{r}\right)^{x}}{1-\frac{g}{r}}\right]^{(r/g)-1}\frac{J_o+\left(\frac{g}{r}\right)^{x}(H_o+K_o)}{J_o+\frac{g}{r}(H_o+K_o)}$$

For cases of J = 0, L^*_{min} is the optimal patent term and Y(x) reduces to:

(8a)
$$Y(x)=\left[\frac{\frac{r}{g}-\left(\frac{r}{g}\right)^{1-x}}{\frac{r}{g}-1}\right]^{(r/g)-1}\left(\frac{r}{g}\right)^{1-x}$$

This expression can be used to generate the values in Table 2 and the curves in Figure 3.

Appendix 3: Effect of the minimum optimal patent term for positive spillover inventions.

Define a function U = W(L*) / W(L^*_{min}), which is the welfare gain associated with an optimal term L*, with the welfare gain at the minimum optimal term L^*_{min} normalized to 1. Where spillovers are not great — i.e., $J_o < g(H_o + K_o)/(r-g)$ — then L* is less than ∞ and U can be found by substituting equations (10) and (11) into equation (4a) above:

(9a)
$$U=\frac{\frac{1}{r-g}\frac{g}{r}\left(\frac{H_o}{Ir(K_o+H_o)}\right)^{(r-g)/g}\left(H_o+K_o+J_o\right)^{r/g}}{\left(\frac{H_o}{Ir}\right)^{(r-g)/g}\left[J_o+(g/r)(H_o+K_o)\right]/(r-g)}$$

Rearranging terms gives:

(10a)
$$U=\left(\frac{J_o}{(K_o+H_o)}+1\right)^{r/g}\frac{1}{\frac{rJ_o}{g(K_o+H_o)}+1}\quad \text{for } L^*<\infty.$$

If r/g ≡w and $J_o / (H_o + K_o)$ ≡M (the spillover ratio), then U can then be expressed as:

(13a)
$$U=\frac{(1+M)^w}{1+wM} \quad \text{for } L^* < \infty.$$

For spillovers are large enough that L* = ∞, then U is the following:

(14a)
$$U=\frac{\frac{J_o}{r-g}\left(\frac{H_o}{I(r-g)}\right)^{(r-g)/g}}{\left(\frac{H_o}{Ir}\right)^{(r-g)/g}\left[J_o+(g/r)(H_o+K_o)\right]/(r-g)}$$

Rearranging terms and substituting M and w as defined above, U can be written as:

(15a)
$$U=\frac{\left(\frac{1}{(1-1/w)}\right)^{w-1}}{\left[1+\frac{1}{wM}\right]} \quad \text{for } L^* = \infty.$$

<u>Small to Moderate Spillovers:</u> Assuming that ∂U/∂M > 0 for all M > 0 (see proof in Appendix 4 below), the limit of U for all inventions with spillovers smaller than the amount needed to produce an infinite optimal patent term (i.e., for all M < g/r-g) will be:

(16a)
$$U(M,w)<\frac{\left(\frac{1}{(1-1/w)}\right)^{w-1}}{\left[1+\frac{r-g}{wg}\right]} \quad \text{for all } M < g/(r-g)$$

(17a)
$$U<\frac{\left(\frac{1}{(1-1/w)}\right)^{w-1}}{\left[2-\frac{1}{w}\right]}\equiv U_{\text{lim1}}$$

U_{lim1} is an upper bound for the U an invention having spillovers smaller than necessary to justify an infinite optimal patent term. Table 3 and Figure 4 show $1/U_{\text{lim1}}$ for various values of r/g.

Large Spillovers: For inventions having spillovers large enough to justify an infinite patent term, the function $U(M, w)$ is given by equation (15a). Assuming that $\partial U/\partial M$ is always positive for positive M, $U(M, w)$ will reach a maximum as $M \to \infty$. Taking the limit of equation (15a), we have:

(18a) $$U \to U_{\lim 2} = \left(\frac{1}{(1-1/w)}\right)^{w-1} \quad \text{as } M \to \infty$$

Table 4 and Figure 5 in the text show $1/U_{\lim 2}$ for various values of r/g.

Appendix 4: Proof that U always increases as M increases.

With the above expressions for U in (13a) and (15a), we can show that U always increases as M increases — i.e., $\partial U/\partial M > 0$ for all $M > 0$. For finite values of L*, $\partial U/\partial M$ is:

$$\frac{\partial U}{\partial M} = \frac{w(M+1)^{w-1}}{wM+1} - \frac{w(M+1)^{w}}{(wM+1)^2}$$

$$\frac{\partial U}{\partial M} = \frac{w(M+1)^{w}}{wM+1}\left[\frac{1}{M+1} - \frac{1}{wM+1}\right]$$

Since M and w are both positive, $\partial U/\partial M$ will be positive if the quantity in brackets is positive — which will be true provided that $w = r/g > 1$. By assumption $r > g$, and so U is always increasing as M increases.

Where $M \geq g/(r-g)$ and thus L* is infinite, $\partial U/\partial M$ is obviously positive for all $w > 1$:

$$\frac{\partial U}{\partial M} = \left(\frac{1}{(1-1/w)}\right)^{w-1} w(1+wM)^{-2} > 0 \quad \text{for all } M > 0 \text{ and all } w > 1.$$

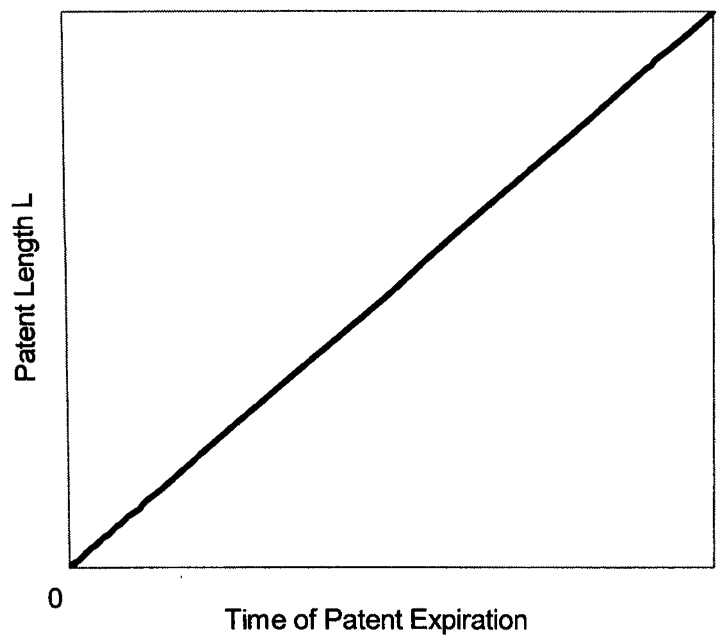

Figure 1: **Increasing patent length (y-axis) always pushes the time of patent expiration later under the assumptions of the Nordhaus model.**

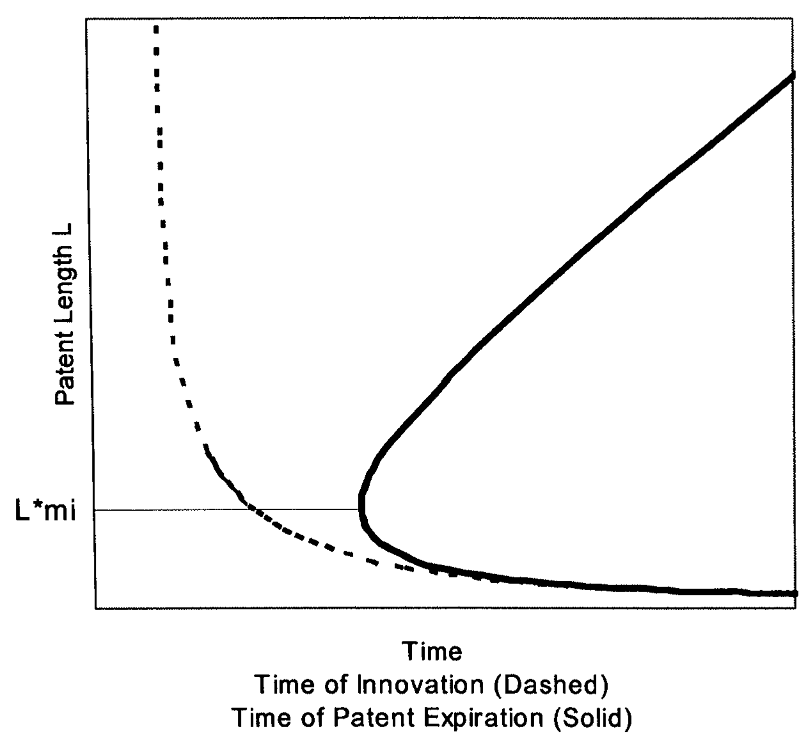

Figure 2: **Where the time of innovation is allowed to vary, increasing patent term up to L^*_{min} moves the time of patent expiration earlier.**

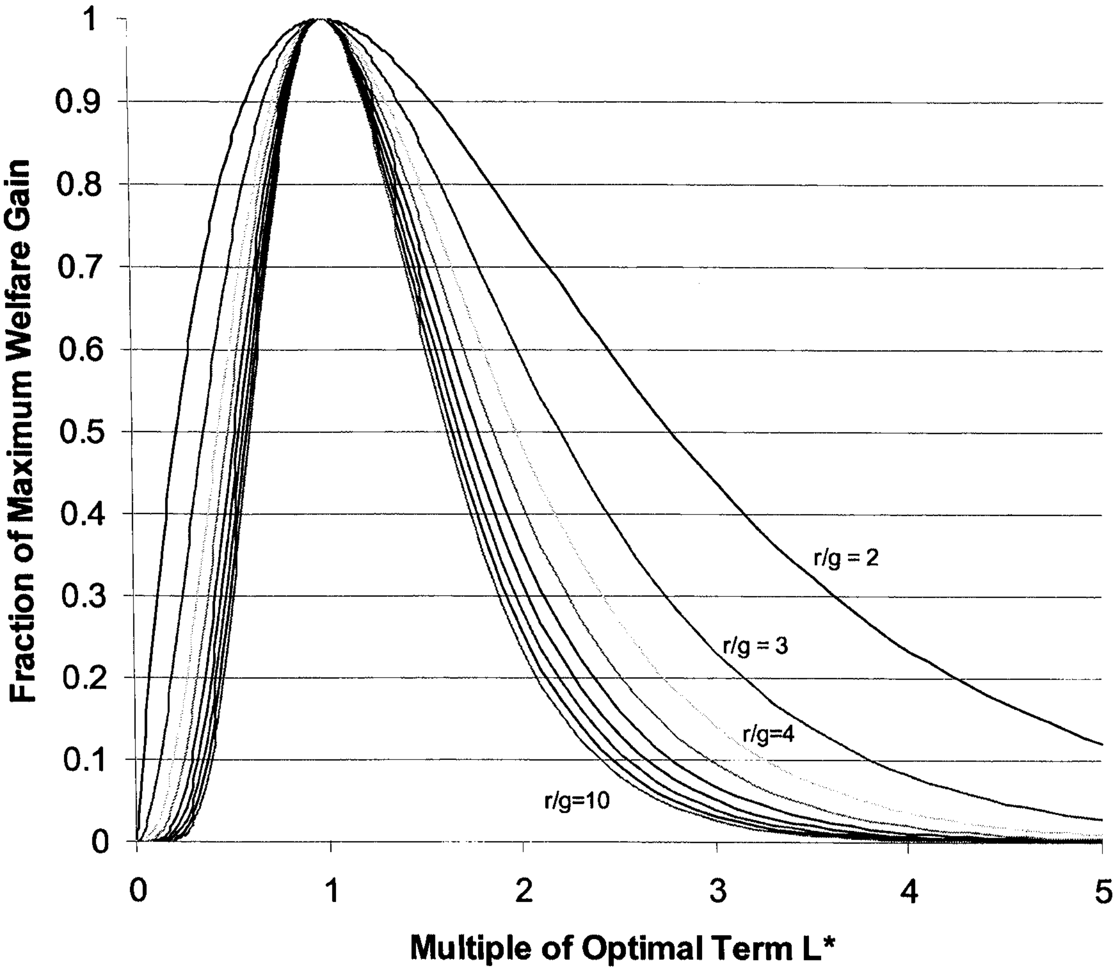

Figure 3: Welfare gain as a function of patent term length for an invention with zero spillovers (J = 0). Maximum welfare gain (L = optimal term L*min) is normalized to 1.

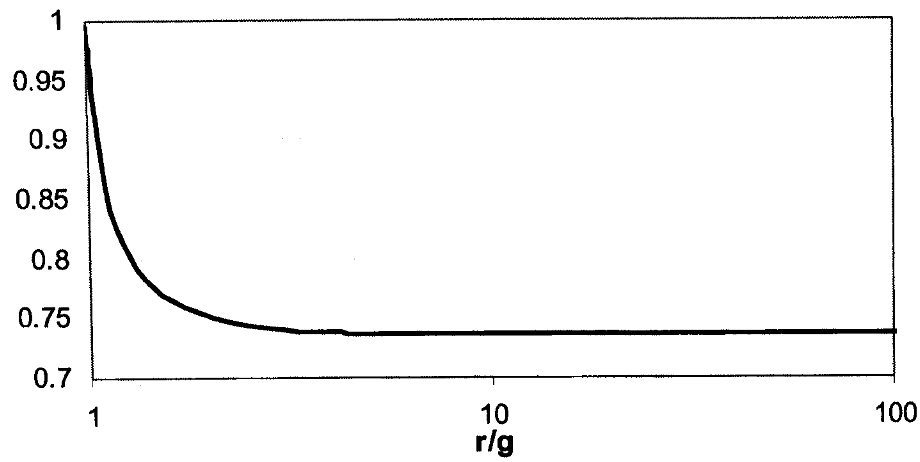

Figure 4: Welfare gain achieved by a patent term of L*min for an invention having spillovers substantial enough to justify an infinite optimal term (i.e., $J_0 = (H_0 + K_0)g/(r-g)$.

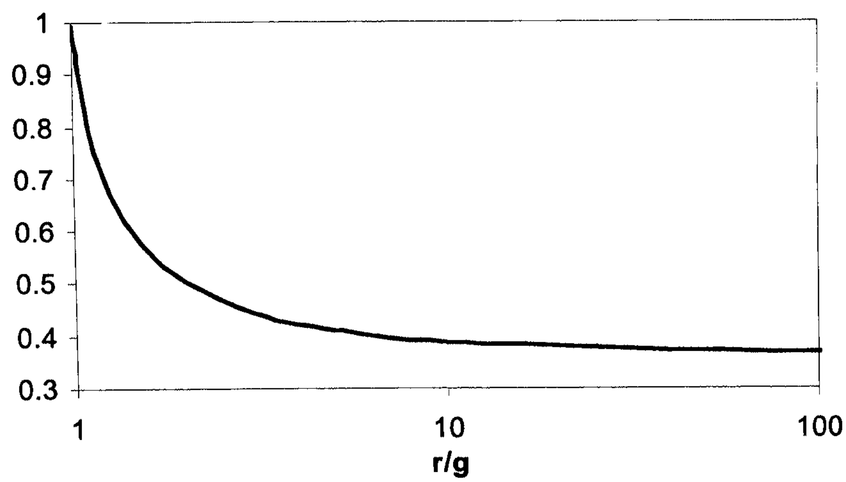

Figure 5: Welfare gain achieved by a patent term of L*min for an invention having very large spillovers (i.e., $J_0 / (H_0 + K_0) \rightarrow \infty$).

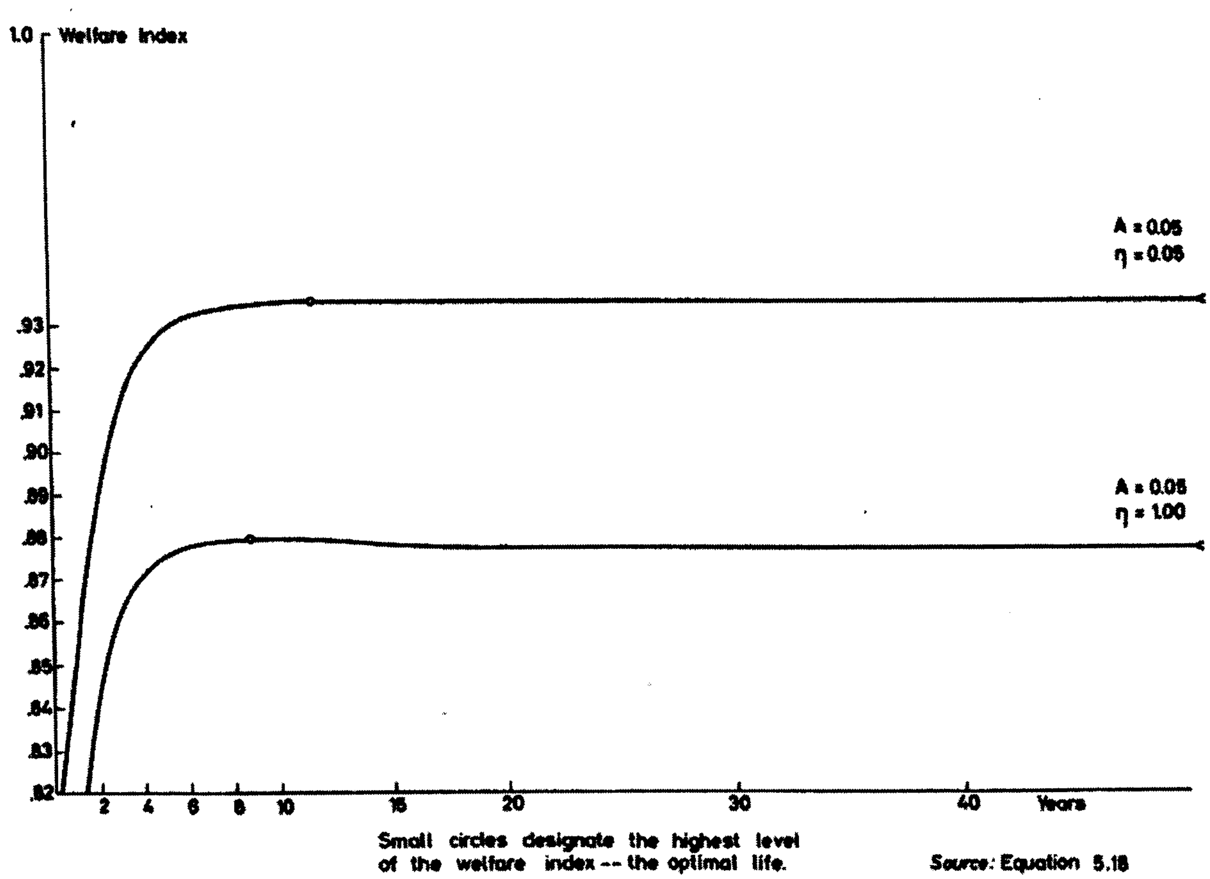

Figure 5.6 Welfare index as function of life of patent

Figure 6: Welfare effects of increasing patent life under the Nordhaus model. (From Nordhaus, 1969).

								Discount Rate (r)						
			.2	0.12	0.11	0.1	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02
	0.09		7.259	9.589	10.03	10.54	NA	NA	NA	NA	NA	NA	NA	NA
	0.08		7.636	10.14	10.62	11.16	11.78	NA	NA	NA	NA	NA	NA	NA
	0.07		8.076	10.78	11.3	11.89	12.57	13.35	NA	NA	NA	NA	NA	NA
	0.06		8.6	11.55	12.12	12.77	13.52	14.38	15.42	NA	NA	NA	NA	NA
growth	0.05		9.242	12.51	13.14	13.86	14.69	15.67	16.82	18.23	NA	NA	NA	NA
rate (g)	0.04		10.06	13.73	14.45	15.27	16.22	17.33	18.65	20.27	22.31	NA	NA	NA
	0.03		11.16	15.4	16.24	17.2	18.31	19.62	21.18	23.1	25.54	28.77	NA	NA
	0.02		12.79	17.92	18.94	20.12	21.49	23.1	25.06	27.47	30.54	34.66	40.55	NA
	0.01		15.77	22.59	23.98	25.58	27.47	29.71	32.43	35.84	40.24	46.21	54.93	69.31

Table 1: Minimum Optimal Term (L^{*}_{min}) for Selected Values of r and g (r > g).

				r/g		
		2	3	4	5	6
	0.5	0.828427	0.696152	0.592593	0.509747	0.442262
	0.6	0.897915	0.813614	0.743277	0.683512	0.631939
	0.7	0.946572	0.900571	0.860712	0.82562	0.794302
multiple	0.8	0.977889	0.958422	0.941221	0.925802	0.911802
of optimal	0.9	0.994849	0.990278	0.986215	0.982553	0.979209
patent	1	1	1	1	1	1
term	1.1	0.995515	0.9916	0.988175	0.985132	0.982389
	1.2	0.983243	0.968898	0.956578	0.945802	0.936224
	1.3	0.964751	0.935339	0.910654	0.889491	0.871007
	1.4	0.941367	0.893916	0.855148	0.82267	0.794866
	1.5	0.914214	0.847151	0.793981	0.750558	0.71419
	2	0.75	0.592593	0.488281	0.41472	0.360232
	3	0.4375	0.23182	0.14131	0.09457	0.06753

Table 2: Welfare gain as a function of patent term length and r/g for an invention with zero spillovers (J = 0). Maximum welfare gain (L = optimal term = L*min) is normalized to 1.

r/g	1.001	1.01	1.1	1.25	1.5	2	5	10	1000
Welfare Gain (L^*_{min})	.994	.964	.858	.802	.770	.750	.737	.7361	.7358

Table 3: Welfare gain achieved by a patent term of L*min for an invention having spillovers substantial enough to justify an infinite optimal term (i.e., $J_0 = (H_0 + K_0)g/(r\text{-}g)$. Welfare gain is expressed as a fraction of the theoretically possible welfare gain from the invention in a patent system with an infinite term.

r/g	1.01	1.1	1.25	1.5	2	5	10	100	∞
Welfare Gain (L^*_{min})	.955	.787	.669	.577	.500	.410	.387	.3697	1/e

Table 4: Welfare gain achieved by a patent term of L*min for an invention having very large spillovers (i.e., $J_0 / (H_0 + K_0) \rightarrow \infty$). Welfare gain is expressed as a fraction of the theoretically possible gain from the invention in a patent system with an infinite term.

*.George Washington University; visiting, spring 2005, New York University. Copyright ©2004. Thanks to Vincenzo Denicolo, Richard Hynes, Eric Kades, Robert Merges, and Robert Stoner for comments.

1.See, e.g., Scherer (1972), at 424 ("To find the socially optimal patent life, one must balance the marginal deferrals of [the deadweight loss triangle] and the (rising) RD costs against the increasing amount of cost reduction . . . stimulated by longer patent lives"); McFetridge and Rafiquzzaman (1986), at 95 (claiming that "[a]n increase in the patent term ... postpones the date at which [the consumer surplus rectangle and the deadweight loss triangle] are realized as surplus."); DeBrock (1985) ("Unfortunately, extension of patent protection by definition brings with it the social inefficiencies recognized in a monopolistic market.").

2. For purposes of this paper, the concept of "patent term" should be taken to mean any period of market exclusivity however obtained. Thus, a "first mover advantage" conferring effective market exclusivity for a period of time should be considered equivalent to a de facto patent term for that period of time.

3. The alternative — that $\partial t_E(L)/\partial L \leq 0$ for all positive L — can be eliminated as a theoretical impossibility because it would require that $t_I(L) \to -\infty$ as $L \to \infty$.

4. Following Nordhaus, others have also assumed that "[s]horter patent lives allow [the gain accruing at patent expiration] to be realized earlier," Kaufer (1989), at 36, or, equivalently, that increasing patent life always makes consumers "worse off because the inventor's monopoly lasts longer," Takalo (2001), at 36.

5. To see this point, note that the basic over-racing problem arises where an investment in innovation (I) occurs at a time when immediate cost of the investment (I*r) is greater than the immediate flow of social returns from the innovation — i.e., where $Ir > SB(t_I)$. This relation cannot occur if

$dt_E(L)/dL \leq 0$ and the following three assumptions are true:

(i) The private cost of innovation is equal to the social cost;

(ii) The flow of private benefits associated with a patent on the innovation, H(t), are equal to or less than the flow of social benefits from the innovation ($H(t) \leq SB(t)$) for all time.

(iii) Competition always dissipates private profits from the innovation ($\pi_i = 0$).

Where increasing patent term is decreasing the time of patent expiration ($dt_E(L)/dL \leq 0$), then the change in private profits due to a marginal increase in patent term will include one term representing additional royalties reaped by earlier innovation, one term representing royalties lost due to earlier patent expiration, and one term representing the cost of investing earlier:

$$\Delta\pi_i = H(t_I)\Delta t_I\, e^{-r t_I} - H(t_E)\Delta t_E\, e^{-r t_E} - Ir\, \Delta t_I\, e^{-r t_I}$$

By assumption (iii) above, we know that $\Delta\pi_i = 0$. Because $dt_E(L)/dL \leq 0$, the term representing the royalties lost due to earlier patent expiration ($H(t_E)\Delta t_E$) must be greater than or equal to zero. Thus, the following holds:

$$Ir \leq H(t_I)$$

Assumption (ii) requires that $H(t_I) \leq SB(t_I)$, and so $Ir \leq SB(t_I)$ and overracing cannot occur. The intuition behind this result is straightforward. Because the patentee has to pay the full social cost of earlier innovation, the patentee must recoup that cost in royalties. If the patentee has advanced the time of innovation to the point where the royalties realized by incrementally earlier patenting do not cover the incremental costs of earlier innovation, then the deficit must be satisfied by capturing additional royalties at the end of the patent term, and that can occur only if the time of patent expiration is moving incrementally later.

6.The assumption is based on a von Stackelberg model in which potential entrants can make R&D

commitments sequentially. See Dasgupta and Stiglitz (1980b), at 10.

7. See also Dasgupta and Stiglitz (1980a, at p. 273), which concludes that the optimum R&D expenditure always increases as the size of the market increases.

8.Throughout this paper, it is assumed that H, J and K are equal to or greater than zero. The possibility that $J < 0$ — which could occur where the invention produces negative externalities — is not investigated, though that situation would obviously decrease the optimal patent term.

9. Romano (1990) also investigated the optimal patent term using a deterministic model that permits the time of innovation, but not of investment, to vary. He also concluded that, for "nondrastic discoveries" (i.e., inventions with zero spillovers during the patent term), "[t]he objective [in finding the optimal patent term] is simply to minimize the date at which the patent will expire and competitive production begins." Id. at 37.

10. The expression for an optimal patent term in equation (11) reduces to the expression derived by Kitti provided that K is set equal to 0. See Kitti (1973) at 27-28. (The variable m in Kitti's notation is equal to $H/(H + J)$ in the notation of this paper.) Kitti assumed that deadweight loss is equal to zero, see Kitti at 23-24 & 65-66, and thus did not reach any conclusions about the effect of deadweight loss on optimal patent term.

11. The welfare calculations discussed in this section of the paper are based on the deterministic model. The appendix supplies the necessary derivations.

12. See also Dore et al. (1993), which note the short optimal patent term produced if the Nordhaus model is modified to assume rivalry for the patent right.

13. Scherer also correctly recognizes that, under the Nordhaus model, longer patent terms produce only increments of innovation with "relatively low benefit-cost ratios — those which in any event are not likely to have a great impact on social welfare." Scherer, at 426. Scherer did not realize that this effect was directly attributable to Nordhaus's assumption about time.

14. Nordhaus limited his inquiry to "run-of-the-mill" process innovations — i.e., innovations that reduce production costs but cause no change in demand during the patent life. The value of the patent on the innovation is then directly proportional to the savings in production costs.

15. A paper by Dore, et al., (1993) also concludes that, under the Nordhaus assumptions, "it is hard to justify a patent for more than 2 or 3 years" where there are decreasing returns to R&D investment. Id. at 19.

16. To see this point, consider an invention with H = a, J = 0, and K = 0. From point (1) in the text, we know that the optimal patent term for this invention minimizes the date of patent expiration; let this patent term be denoted as $Term_1$ (J=0). Now consider an invention having H = a, J = 0, and K = c, where c is arbitrarily large. From footnote 10 in Denicolo, we know that the optimal patent term is unaffected; it remains $Term_1$ (J=0), which is not a function of K. Now consider an invention with H = a, J = b, and K = c. From Denicolo's conclusions, we know that the optimal patent term for this invention will be greater than $Term_1$ (J=0); let the increment to patent term be defined as $Term_2$, which is always positive and is a function of at least J. Thus, we know that, for any invention of H = a, the optimal patent term will be in the form:

$$\text{Optimal Term} = Term_1\,(J{=}0) + Term_2\,(J).$$

Denicolo's results also show that the optimal patent term is decreasing in K where $J \neq 0$. Since $Term_1$ is not a function of K, $Term_2$ must be. Thus, the optimal patent term must be in the form

$$\text{Optimal Term} = Term_1\,(J{=}0) + Term_2(J, K)$$

where $Term_2$ is always positive and is a function of J and K.

The work of Romano (1990) can also be used to show that the optimal patent term in J = 0 cases is also a minimum optimal term when J, K > 0. Again, Romano does not explicitly state this result, but it can be easily derived from equations (3) and (4) in his paper, plus the observation (which is explicitly stated) that the optimal term in the zero spillover case does not depend on deadweight loss (K).

17.Deadweight loss could be considered a social cost of a patent if (as in Wright 1983) the patent system is being compared to a system where the invention is socially subsidized in some fashion and, once created, the invention is treated as a public good. In that case, the social cost of the patent system would include the deadweight loss created by the patent right and would properly be compared to the costs of social subsidization, including, among other things, the tax distortions necessary to raise the revenue needed to pay the subsidies. Duffy (2004a). However, the literature on optimal patent term compares patent systems with terms of differing lengths.

18.The U.S. denies patent protection to inventions that do not meet a certain standard of inventiveness or nonobviousness. See 35 U.S.C. § 103. Many other countries provide some patent-like protection for obvious developments, but the term of such protection is shorter than that afforded for patents. See Janis (1999).

19.As discussed by Denicolo (1996), the concept of patent breadth is somewhat ambiguous. While a reduction in patent breadth always corresponds to increasing competition in the product market, the increased competition could result from several different policy changes, including the addition of compulsory licensing, reducing the costs of imitation or permitting certain applications of the innovation to be considered non-infringing.

20.The literature on patent breadth generally holds patent rewards constant and minimizes the deadweight loss. See Klemperer (1990), Gilbert and Shapiro (1990), Gallini (1992) and Wright (1999). Under these models, any reduction in patent breadth must be matched by a corresponding increase in patent life.