OPTIMAL PATENT RENEWALS*

by

Francesca Cornelli
London Business School and CEPR

and

Mark Schankerman
London School of Economics and Political Science, and EBRD

Discussion Paper
No.EI/13
January 1996

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 – 7955 6674

---

* We would like to thank Leonardo Felli, Nancy Gallini, Ed Green, Bronwyn Hall, Paul Klemperer, Ariel Pakes and Jean Tirole for comments. We also thank participants in seminars at LSE, UCL, NBER, Tel Aviv, CREST-LEI, Studienzentrum Gertzensee, Toulouse and the 1995 European Economic Association meetings in Prague. Orli Arav provided excellent research assistance. All errors remain our own. Part of this work was done while Francesca Cornelli was visiting CREST-LEI and Tel Aviv University, whose hospitality is gratefully acknowledged.

Abstract

When firms have different R&D productivities, it may be welfare increasing to differentiate patent lives across inventions. The reason is that any uniform patent life provides excessive incentives to do R&D to the low productivity firms and insufficient incentives to the high productivity firms. Such a differentiated scheme is implementable through renewal fees, which endogenously determine an optimal pattern of patent lives. We characterise the optimal pattern of patent life-spans and show how it depends on key features of the economic environment, such as the degree of heterogeneity in R&D productivity across firms, the ability of patentees to appropriate the potential rents generated by R&D and the learning process about the value of the innovation. We illustrate the potential welfare gains associated with optimal renewal schemes through simulation analysis.

Keywords: patents, renewal fees, R&D, fees, welfare gains, productivity, firms.

Contact address: Dr Mark Schankermann, Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE. Email: m.schankerman@lse.ac.uk

1. Introduction

Most patent systems require that patentees pay annual renewal fees in order to maintain patent protection up to a statutory patent life. Failure to pay the renewal fees permanently cancels patent protection. Despite the fact that these fees are relatively low (though they do vary considerably across countries and rise with patent age), more than half of patents are voluntarily canceled within ten years after the date of patent application. Econometric studies of patent renewal data have quantified the value of property rights embodied in patent protection and have shown clearly that patentees respond to renewal fees in their decision whether to maintain patent protection (Schankerman and Pakes 1986, Pakes 1986, Schankerman 1991 and Lanjouw 1992). Even though all countries impose a uniform statutory patent life, patent renewal fees have effectively created *de facto* differentiation in patent lives.

However, in practice patent renewal fees are used simply as a fiscal device (to fund patent offices) and there is no reason to believe that the variation in effective patent life produced by these fee schedules is welfare-improving. The theoretical literature focuses on the determination of the optimal *uniform* patent life and, more recently, on other dimensions such as the width of the patent (Nordhaus 1969, Klemperer 1990, Gilbert and Shapiro 1990). This paper argues that patent renewal fees can and should be designed strategically by the government in order to increase welfare. We derive conditions that characterize optimal (welfare maximizing) patent renewal fee schedules, which endogenously determine an optimal pattern of patent lives. We show how the optimal pattern of patent life-spans depends on key features of the economic environment, including the effectiveness of R&D (in generating product quality improvements or cost reduction), the degree of heterogeneity in R&D productivity across firms, and the ability of patentees to appropriate the potential rents generated by R&D. These features may be expected to differ across industries, so that the paper provides analytical grounds for differentiating optimal patent renewal schemes across industries.

The analysis is based on the observation that variations across firms in their R&D productivity imply that it may be welfare increasing to differentiate patent lives across firms (i.e. inventions). The reason is that any uniform patent life provides excessive incentives to do R&D to the low productivity firms and insufficient incentives to high productivity firms. The consequence is both a socially suboptimal level of R&D and distribution of R&D across firms. The latter, in turn, implies that there are excessive social costs of inducing R&D activity (i.e., the deadweight loss from patent protection can be lowered for any given level of R&D).

The theoretical problem, and policy challenge, is to find an instrument which can be used to differentiate patent lives optimally. This problem must be addressed in the context of asymmetric information, since the use of patents as a policy instrument only makes sense if the government has incomplete information about the cost and value of inventions (Wright 1983). This asymmetry of information requires an implementable revelation mechanism. The key observation in this paper is that renewal fees can be used precisely for this purpose.

The government derives the optimal patent life conditional on the firm's R&D productivity parameter (taking into account the required R&D incentive). The government does not observe the firm's productivity parameter and sets the incentive compatible fee schedule in order to implement this differentiated patent-life scheme. This is equivalent to the government offering a menu of patent life-spans and associated fees so that a firm, when applying for a patent, selects its preferred life-span/fee combination. Alternatively, the government can offer an annual renewal fee schedule and firms make a sequence of renewal decisions. If firms face no uncertainty about the profitability of their inventions, once they have been produced, these two indirect mechanisms are equivalent. However if there is post-invention uncertainty (i.e. private information is only about expected profits), then we show that the annual renewal scheme is the superior mechanism.

We use simulation analysis to show how the optimal patent mechanism can be easily implemented and to illustrate the potential welfare gains associated with optimal renewal schemes. The welfare implications of the optimal renewal scheme are compared (under various parameterizations) to an "optimal" uniform patent life without renewal fees. We show that the optimal patent length for "high" R&D productivity firms is considerably longer than existing statutory patent lives, that optimal renewal fees should rise much more sharply with patent length than the existing statutory fee schedules, and that the optimal patent mechanism can yield significant welfare gains.

We introduce the model in Section 2. Section 3 derives the optimal uniform patent life, while Section 4 considers the simpler case in which firms are not *ex ante* heterogeneous. The main result is presented in Section 5: the optimal scheme, with differentiated patent lives, when firms have different R&D productivity. Section 6 uses simulation analysis to characterise the optimal patent mechanism and to illustrate the associated welfare gains. Section 7 and 8 extend the analysis to the case in which firms learn more about the value of the invention after obtaining the patent and the case in which firms differ in their ability to appropriate the rents from patents. Section 9 provides concluding remarks and directions for future research.

2. The model

The timing of the model is the following: at the beginning of period 0 firms invest in R&D, which yields an invention at the end of the period. At the beginning of period 1 firms may apply for a patent of length T. After they obtain the patent, there is an infinite number of periods in which the innovation is used (protected by the patent or not, since the patent will eventually expire).

It is important to characterize precisely the information structure of the problem because, as we will show, different sources of information affect the design of patent policy in very different ways. We distinguish between *heterogeneity*, which refers to firm-specific characteristics that are private information, and *uncertainty* which refers to the stochastic elements in the production and marketing of inventions which are known only to the firm after their realization. *Ex ante* (pre-invention), firms are heterogeneous in terms of their productivity. *Ex post*, the value of inventions will be heterogeneous both because of *ex ante* heterogeneity and the stochastic R&D outcome. We will show that *ex ante* heterogeneity makes it optimal and feasible for the government to differentiate patent lengths. However, this result does not hold if the *ex post* heterogeneity arises only from the stochastic R&D process.[1]

We let μ denote the "size" of the innovation. This could parameterize the cost of producing a good, in the case of a process invention, or the level or elasticity of demand, in the case of a product invention. For simplicity, we analyze the case in which μ affects the level of demand: given μ the demand in each period is given by $Q(p \mid \mu)$, with $\frac{\partial Q}{\partial \mu} > 0$. The analysis requires specification of the appropriability environment in which the patent holder operates. For presentational simplicity we assume that during the patent lifetime the innovating firm is a monopolist charging a uniform price and that competitive pricing prevails after patent expiration. The analysis which follows also holds for more sophisticated forms of appropriation, such as price discrimination and licensing of the invention, but the functions would be more complicated. A more general treatment of appropriability is given in Section 8.

Assuming for simplicity that the marginal production cost is zero, the firm with a patent sets p in each period in order to maximize

[1] There is an additional source of (*ex post*) uncertainty which arises when the firm learns about the value of its invention after the patent decision. This issue is explored in Section 7.

$pQ(p \mid \mu)$. Define

$$p^*(\mu) \equiv \underset{p}{\operatorname{argmax}}\, pQ(p \mid \mu)$$

Then the profits per period during the patent life are given by

$$\pi(\mu) \equiv p^*(\mu)Q(p^*(\mu) \mid \mu) \tag{1}$$

The total welfare per period (profits plus consumer surplus) during the patent lifetime is given by

$$\hat{w}(\mu) \equiv \int_{p^*(\mu)}^{\infty} Q(p \mid \mu)dp + \pi(\mu) \tag{2}$$

The profits and total welfare per period after the patent is expired will depend on the market structure that arises thereafter. We assume that the market is perfectly competitive after the patent expires, so that profits are zero and welfare is given by:

$$\hat{B}(\mu) \equiv \int_{0}^{\infty} Q(p \mid \mu)dp \tag{3}$$

The total welfare maximized by the government is therefore given by

$$\hat{W}(T, \mu) \equiv \int_{0}^{T} \hat{w}(\mu)e^{-rt}dt + \int_{T}^{+\infty} \hat{B}(\mu)e^{-rt}dt \tag{4}$$

It is easily confirmed that the function $\pi(\mu)$ given in equation (1) is monotonic increasing. By the envelope theorem, $\frac{d\pi}{d\mu} = p^*(\mu)\frac{\partial Q}{\partial \mu} > 0$. This means we can invert the function and express the size of the invention as a function of the level of (maximized) profits as follows:

$$\mu = g(\pi) \equiv \pi^{-1}(\pi)$$

Substituting for μ we can write

$$w(\pi) \equiv \hat{w}(g(\pi))$$

$$B(\pi) \equiv \hat{B}(g(\pi))$$

$$W(T,\pi) \equiv \hat{W}(T, g(\pi))$$

where it is important to recall that π refers to the profits per period the innovating firm earns during the patent lifetime. The total profits maximized by the firm are given by

$$\Pi(\pi, T) \equiv \int_0^T \pi e^{-rt} dt = \frac{\pi}{r}(1 - e^{-rT}) \tag{5}$$

The assumptions that the market structure is perfectly competitive after patent expiration and that the marginal cost is zero can be relaxed, but the functions would be considerably more complex. What matters for the analysis is that it is possible to express welfare as a monotonic function of π.

Notice that we can write

$$B(\pi) = w(\pi) + D(\pi)$$

where

$$D(\pi) \equiv \int_0^{p^*(\pi)} Q(p \mid \mu) dp - \pi \tag{6}$$

is the deadweight loss. It can be easily verified that $\frac{dD(\pi)}{d\pi}$ cannot be signed, so "larger" inventions may be associated with either higher or lower deadweight loss. However, we assume that "larger" inventions do generate higher social benefits after patent expiration - i.e.,

$\frac{dB}{d\pi}(\pi) > 0$.

Using these expressions we write the government's objective function as

$$W(T,\pi) = \int_0^{+\infty} w(\pi)e^{-rt}dt + \int_T^{+\infty}[B(\pi) - w(\pi)]e^{-rt}dt = \frac{w(\pi)}{r} + \frac{D(\pi)}{r}e^{-rT} \quad (7)$$

The interpretation of the objective function is as follows: the first term represents the present value of consumer surplus and profits which accrue each period (after patent expiration, profits accrue as surplus). The second term is the additional gain (formerly deadweight loss) which begins from period T, when the patent expires.

The next step is to describe the R&D process generating inventions. In what follows we refer to the profit π as the output of the R&D activity. In fact R&D produces an invention of size μ, from which the maximum profits π are uniquely determined. Hence π is a summary statistic for the product of R&D, given the appropriability environment.

We characterise the R&D process in the following way:

$$\pi = \theta z \quad (8)$$

where z is the R&D input, which could be effort or the amount of R&D resources, and θ is the marginal productivity of such input. The parameter θ summarizes two stages of the innovation process: the ability of the firm to produce inventions (R&D productivity) and the ability of the firm to appropriate returns from its patented invention (market efficiency). For brevity we will refer to θ as the R&D parameter. A firm characterised by a high θ is more efficient in the R&D process.

In general both stages of the R&D process are stochastic, but for simplicity we assume that the firm has perfect control in the choice of the size of the innovation, and hence in π, and knows its R&D parameter. We show later that this assumption does not involve any loss of generality. However, the government does not observe the value of the invention π. Since we are assuming that the R&D process is deterministic, this implies that the government does not know θ, but it is assumed to know that θ is drawn from the distribution function $G(\cdot)$, defined over the interval $[0, \bar{\theta}]$, with density $g(\cdot)$.

Finally, the disutility of effort z (or the cost of R&D resources) is given by a function $\psi(z)$ such that $\psi' > 0$ and $\psi'' > 0$.[2]

3. The optimal length of a patent

If the government knew θ it could compute the first-best level of effort. The first best level of effort is defined by:

$$z^{**}(\theta) = \underset{z}{\operatorname{argmax}}\left\{\frac{B(\theta z)}{r} - \psi(z)\right\} \tag{9}$$

However, the first best is not achievable because the government

[2] The specification used here involves linear profits and convex costs in z. Alternatively, we could use a concave "R&D production function" minus resource costs. The latter specification is technically more difficult to handle.

cannot observe the firm's effort and therefore must provide incentives to induce R&D by issuing patents. Let us begin by assuming that the government is limited to setting the optimal uniform length of a patent. Given a patent of length T, the firm sets its R&D to maximize expected net profits

$$\Pi(\theta z, T) = \int_0^T \theta z e^{-rt} dt - \psi(z)$$

which yields

$$z^*(\theta, T) = \psi'^{-1}\left(\frac{\theta}{r}(1 - e^{-rT})\right) \tag{10}$$

Notice that z^* increases as T rises. The socially optimal T is given by

$$\max_T E_\theta \left\{ \left[\frac{w(\pi)}{r} + D(\pi)\frac{1}{r}e^{-rT} - \psi(z) \right] \right\} \tag{11}$$

s.t.

$$\pi = \theta\psi'^{-1}\left(\frac{\theta}{r}(1 - e^{-rT})\right) \tag{12}$$

Substituting equation (12) in (11), the first order condition yields:

$$E_\theta \left\{ \frac{\theta^2}{r}\frac{\partial \psi'^{-1}}{\partial z}\left[\frac{\partial w}{\partial \pi} + e^{-rT}\frac{\partial D}{\partial \pi}\right] \right\} = E_\theta \left\{ \psi'\theta\frac{\partial \psi'^{-1}}{\partial z} + D(\pi) \right\}. \tag{13}$$

Determination of the optimal (uniform) patent life involves equating the marginal social benefit and cost of extending T. To highlight this intuition, we multiply both sides of the first order condition by e^{-rT} and rewrite it as follows:

$$E_\theta \left[\frac{dB}{d\pi}\frac{d\pi}{dT} \right] = E_\theta \left[\frac{\partial \psi}{\partial z}\frac{dz}{dT} + D(\pi)e^{-rT} \right]$$

The left hand side is the expected marginal benefit from extending the patent life T. This reflects both the incentive effect on R&D and hence the size of the invention (profits), and the marginal social valuation of larger inventions. The right hand side is the expected marginal social cost, comprised of the additional resource cost induced by the increase in T and the discounted value of the additional period's deadweight loss created by the patent.

To simplify the analysis, we focus in the rest of this paper on the special case of quadratic R&D costs $\psi(z) = \frac{1}{2}z^2$. The results would carry over to the more general specification. Even with quadratic costs, in general there is no closed form solution for the optimal patent life. A very simple benchmark case which does yield a closed form is the linear specification of welfare and deadweight loss: $w(\pi) = \alpha\pi$ and $D(\pi) = \beta\pi$. In this case the optimal length is

$$T_U = \frac{1}{r}[\ln(1 + 2\beta) - \ln(1 + \beta - \alpha)] \tag{14}$$

Note that in this special case the optimal (uniform) patent life is independent of the distribution of the productivity parameter θ. This does not hold in general.[3]

It is easy to verify that all firms find it optimal to do at least some R&D as long as $\theta > 0$ and $T_U > 0$. This is so because, with a deterministic R&D process, a firm can always do very little R&D and

[3] The linear specification requires the additional restriction that $\beta > \alpha - 1$. This means that the (constant) marginal loss in deadweight loss must exceed the gain in consumer surplus per period from an increase in patent length. If the equality holds, the socially optimal (uniform) patent life is infinite.

obtain a small invention. Since the function $\psi(\cdot)$ is convex it will not be costly to do a very small amount of R&D. However, the challenge for the government is to design a mechanism which both induces the optimal overall amount of R&D and the optimal distribution of R&D across firms which are heterogeneous in terms of the productivity parameter θ. We turn next to this issue.

4. Can renewal fees screen?

In this section we consider the conditions under which it is feasible to use patent renewals to screen among different firms (inventions). The standard approach in the literature is to model optimal patent policy under the constraint that firms need to obtain a specified level of profit to compensate for the R&D cost. This approach indirectly takes into consideration the issue of R&D incentives, but it does not recognize heterogeneity across firms in R&D productivity. In specifying the same required compensation, it is implicitly assumed that all firms have the same value of θ. In other words, the standard approach only allows for *ex post* heterogeneity across firms. All the firms are identical *ex ante*, and it is only because of the stochastic nature of the R&D process that firms will obtain inventions of different sizes. An equivalent interpretation is that the government knows the level of R&D productivity of a firm, but not the actual product of the R&D process.

We begin the analysis from this point of view. In such a case, the government faces a number of firms with inventions of different values π and private information about those values.[4] We examine whether it is feasible and socially desirable to screen (i.e., differentiate patent length) among different inventions. The government can offer a schedule $\{T, f\}$, where T is the length of the patent and f is the fee paid, in order to discriminate among different innovations. The uncertainty about the values π is modeled by assuming that π is distributed according to the distribution $G(\pi)$ in the interval $[0, \bar{\pi}]$. The optimal schedule would be defined by the problem

$$\max_{T,f} \int_0^{\bar{\pi}} \left[\frac{W(\pi)}{r} + \frac{D(\pi)}{r} e^{-rT(\pi)} \right] dG(\pi) \tag{15}$$

s.t.[5]

$$\int_0^{T(\pi)} \pi e^{-rt} dt - f(T(\pi)) = \frac{\pi}{r}(1 - e^{-rT(\pi)}) - f(T(\pi)) \geq V, \quad \forall \pi \tag{16}$$

and

$$\pi = \operatorname*{argmax}_{\hat{\pi}} \frac{\pi}{r}(1 - e^{-rT(\hat{\pi})}) - f(T(\hat{\pi})), \quad \forall \pi, \forall \hat{\pi} \tag{17}$$

Equation (16) is the individual rationality constraint which guarantees that firms actually undertake R&D. Equation (17) is the incentive compatibility constraint which ensures that the government is able to screen. From the usual transformation (see Myerson 1981) we find

---

[4] An equivalent interpretation is that the government faces one firm with an invention whose value π is unknown.

[5] Note that we are considering the individual rationality constraint for each *ex post* π, and not on average. From one perspective, the expected utility over all π is relevant since that is what it matters when firms are investing in R&D. However, when the firm applies for the patent, the *ex post* individual rationality constraint must be introduced because the firm must be guaranteed enough profits in order not to bypass the patent (and protect it by second best means). For simplicity we follow the literature in focusing on the *ex post* individual rationality constraints.

that the fee schedule which guarantees incentive compatibility is

$$f(\pi) = \frac{\pi}{r}(1 - e^{-rT}) - \int_0^{\pi} \frac{s}{r}(1 - e^{-rT})ds - V \tag{18}$$

To satisfy the second order conditions for incentive compatibility it is necessary that

$$\frac{df}{d\pi} \geq 0 \tag{19}$$

Because the level of π is not endogenized here, it is clear that the unconstrained maximization of equation (15) would lead the government to set $T = 0$. More generally, the maximization of (15) subject to individual rationality constraint (16) shows that the optimal fee schedule is decreasing in π—i.e. $(df/d\pi) \leq 0$. Since this violates equation (19), in this case welfare maximization implies that the government chooses a unique patent length and does not discriminate across inventions.

The intuition behind this result is straightforward. in this setting, the government would like to set T as low as possible in order to minimize the deadweight loss. The more valuable is the invention, π, the lower is the T necessary to guarantee minimum profits V. Therefore, welfare maximization requires that the optimal schedule is decreasing with π. However, such a fee schedule can never be incentive compatible, regardless of the level of the fees $f(\pi)$. A firm would always under-represent its true π since it would obtain a longer patent at a lower cost.

This result depends partly on the assumption that patent fees do not enter the government objective function: in other words, fees are treated as transfers among different agents that do not affect social welfare. We can easily relax this assumption by treating proceeds from patent fees as a substitute for public funds that have a shadow cost (as in Laffont and Tirole, 1993). However, for a sufficiently high shadow cost, the government would always find it optimal to discriminate because the distortionary effect of increasing patent lengths would be lower than the use it can make of the funds in other sectors. Therefore, to separate the determination of optimal patent policy from more general public finance issues, we assume that fees do not enter the objective function without any loss of generality.

5. Differentiating patent life-spans

In the previous section, the reason it is not optimal for the government to differentiate patent length is that there is no *ex ante* heterogeneity in the R&D process. If firms do not differ in their ability to perform R&D, there is no incentive effect and no welfare gain from differentiating patent life-spans. When firms do differ, both the overall level and the distribution of R&D across firms are subject to policy influence. In Section 2 we model the R&D process by assuming that firms are characterized by a different parameter θ, which summarizes both R&D productivity and market efficiency.

A larger θ represents higher marginal productivity in the R&D process. If the government recognizes that firms are heterogeneous (even if it does not know the value of θ for each firm), it may want to provide an incentive structure that shifts the distribution of R&D effort toward the high θ (low cost) firms. Note that informational asymmetry about the value of the invention, π, can arise not only from the stochastic nature of the R&D process, but also from incomplete information of the government about the abilities of firms to perform R&D. We believe that this is a realistic characterization of the R&D process. The existing literature typically focuses on *ex post* variation in the value of invention without specifying its source. However, in Section 4 we showed that the source of *ex post* heterogeneity makes an important difference. If it arises only from uncertainty in the R&D process, but firms are equally good in performing R&D (and the government knows how good they are) then there is no point in screening and a uniform patent life is optimal. In this section we show that if firms differ in their R&D productivity (or the government is uncertain about their productivity), then it is optimal for the government to set differentiated patent lengths.

The intuition is simple. With a given T_U, a firm with a higher θ already obtains higher profits during the patent lifetime. But is it sufficient? The answer is no: in setting a uniform patent length, the government averages across firms with different productivity levels, without taking into account the *ex ante* heterogeneity. Thus it provides too few incentives to the more productive firms, and at the same time too much incentive to less productive firms. The consequence is that the social cost of doing R&D is not minimised.

In what follows we focus on *ex ante heterogeneity*. The R&D process is assumed to be deterministic, so that if the government knew θ then it would have perfect information. Of course this is not realistic, but we lose no generality in doing so. In Section 4 we showed that *ex ante uncertainty* will not induce any screening, so for the purpose of finding the optimal schedule we can ignore it.[6]

Let us assume now that the government can use different patent length associated with fees and derive the optimal schedule. In this section we consider a mechanism in which the government offers to the firm the choice among different patent lengths, where each length is associated with a different up-front payment. In the framework used in this section, this mechanism will prove to be equivalent to one in which the firm decides each period whether to renew the patent for a stipulated renewal fee schedule. In the next section we introduce *ex post uncertainty* about the future profits of an invention, and we show that the patent renewal scheme is actually superior to a lump sum payment scheme.

By the Revelation Principle, we can restrict attention to the direct mechanism where the firm announces $\hat{\theta}$ and the revelation mechanism is $\{T(\hat{\theta}), f(\hat{\theta})\}$. The firm facing the this schedule will choose z in order to maximize expected net profits. The privately optimal amount of

[6] Making the R&D process stochastic would not change the argument in favor of differentiating patent lengths, but it would modify the form of the optimal patent schedule. For example, suppose the firm is risk neutral and $\pi = \theta z + \epsilon$. The profit maximizing level of R&D in equation (20), z^*, would be unchanged, but the welfare maximization problem in equation (21) would become

$$\max_{T,f} E_\epsilon \int_0^{\bar{\theta}} \left[\frac{W(\theta z^* + \epsilon)}{r} + \frac{D(\theta z^* + \epsilon)}{r} e^{-rT(\theta)} - \frac{1}{2} z^{*2}\right] dF(\theta).$$

For a given distribution of ϵ, standard simulation techniques could be used to solve this problem. We intend to pursue this line of research in future work.

R&D is given by

$$z^*(\theta,\hat{\theta}) = \frac{\theta}{r}(1 - e^{-rT(\hat{\theta})}) \tag{20}$$

Therefore the welfare maximization problem becomes:

$$\max_{T,f} \int_0^{\bar{\theta}} \left[\frac{W(\theta z^*)}{r} + \frac{D(\theta z^*)}{r} e^{-rT(\theta)} - \frac{1}{2} z^{*2}\right] dF(\theta)^7 \tag{21}$$

subject to

$$U(\theta,\hat{\theta}) = \int_0^{T(\hat{\theta})} \theta z^* e^{-rt} dt - \frac{1}{2}(z^*)^2 - f(\hat{\theta}) \geq 0, \ \forall\theta \tag{22}$$

and

$$\theta = \underset{\hat{\theta}}{\operatorname{argmax}}\, U(\theta,\hat{\theta}), \quad \forall\theta,\hat{\theta} \tag{23}$$

From the usual transformation, we obtain the fee schedule that guarantees incentive compatibility (see Appendix):

$$f(\theta) = \frac{\theta^2}{2r^2}(1 - e^{-rT(\theta)})^2 - \int_0^{\theta} \frac{s}{r^2}(1 - e^{-rT(s)})^2 ds \tag{24}$$

This fee is set equal to the maximized present value of profits (net of R&D costs) minus the information rent which must be left to the firm to induce revelation of θ. Once we have found the optimal schedule of patent lengths, $T(\theta)$, the fee schedule $f(\theta)$ in equation (24) guarantees that it is implementable. The second order condition for incentive compatibility implies that

$$\frac{dT}{d\theta} \geq 0.$$

[7] As in the previous section, we assume that fees do not enter the objective function. However, as we will show later, if fees are negative we may want to introduce them.

Equation (24) is monotonic increasing in θ, but for low values of θ it could take negative values. The intuition is as follows: to ensure that the optimal patent schedule is incentive compatible, the government must increase the patent life-span with θ. If the government does not want patents which are "too long", it could start from a very short length and subsidize the less efficient firms. However, this may be optimal only because we have assumed that fees do not enter the welfare function. We explained earlier why this is the best assumption for positive fees—essentially, we do not want the optimal patent policy to be driven by government revenue considerations. However, when negative fees are allowed, this may be more questionable for two reasons. First, it may then become necessary for the government to ensure that the firms receiving subsidies are actually producing inventions (and such monitoring may be costly). Second, raising public funds may involve a shadow cost because of distortionary taxation. We can easily take the shadow cost of funds into account, either by adding a non-negativity constraint on fees or by incorporating the cost of negative fees directly into the objective function. As a result, it may not be optimal for the government to provide a patent to inventions of less than a minimum size (equivalently, to firms of sufficiently low R&D productivity). In the present context, this can be interpreted as the government imposing a minimum standard for patentability.[8]

[8] If we constrain the fees to be non negative or introduce explicitly the costs of subsidizing into the objective function, then the present framework—which has a non-stochastic R&D process—implies that low productivity firms (with $\theta \leq \theta^*$) will choose not to do R&D since their invention will not generate sufficient profits to cover the minimum patent fee. Since the government can only discriminate on the basis of the outcome of the R&D process, in a more general framework with a stochastic R&D process, some high productivity firms may also produce inventions which do not meet the threshold. However, this will occur

To obtain the socially optimal patent schedule we maximize the objective function (21). The first order conditions have the same form as those in equation (13), except that we now allow T to vary with θ (and specialize for quadratic specification of effort costs):

$$R(T,\theta) \equiv \frac{\theta^2}{r}\left[\frac{\partial w}{\partial \pi} + e^{-rT}\frac{\partial D}{\partial \pi} - (1-e^{-rT(\theta)})\right] - D(T(\theta)) = 0 \quad (25)$$

The interpretation is similar to the one in (13), but now we must check whether the optimal patent schedule is incentive compatible, i.e. $T(\theta)$ is increasing in θ.

From the implicit function theorem we know that $\frac{dT}{d\theta}$ has the same sign as $\frac{\partial R}{\partial \theta}$, which can be written as:

$$\frac{\partial w}{\partial \pi} + (2e^{-rT}-1)\frac{\partial D}{\partial \pi} - (1-e^{-rT}) + \frac{\theta^2}{r}(1-e^{-rT})\left[\frac{\partial^2 w}{\partial \pi^2} + e^{-rT}\frac{\partial^2 D}{\partial \pi^2}\right] \quad (26)$$

Using the first order condition (25) we obtain

$$\frac{r}{\theta^2}D - (1-e^{-rT})\frac{\partial D}{\partial \pi} + \frac{\theta^2}{r}(1-e^{-rT})\left[\frac{\partial^2 w}{\partial \pi^2} + e^{-rT}\frac{\partial^2 D}{\partial \pi^2}\right] > 0 \quad (27)$$

If $T(\theta)$ is increasing in θ the government finds it optimal to give a longer patent to "larger" inventions. In such a case we know that the optimal differentiated patent schedule is implementable, using the corresponding fees given by equation (24). However, if $T(\theta)$ is decreasing in θ, welfare maximization requires that a uniform patent length

less frequently than for low productivity firms (in a stochastic dominance sense) and only firms below the productivity threshold will choose not to undertake R&D. In other words, there will be (second-best) *ex ante* efficiency, but *ex post* inefficiency caused by asymmetric information.

be set. If $T(\theta)$ is non monotonic, then the optimal schedule will have a flat part where $T(\theta)$ is non monotonic.[9]

The sign of the expression in (27) depends on how welfare and deadweight loss vary with profits (size of the invention). Thus the shapes of these functions determine whether it is optimal for the government to differentiate patent lengths. While not surprising, this make clear that optimal patent policy depends heavily on how the private and social benefits of invention are modeled , and this may vary across industries.

If $B''(\pi) > 0$ the bracketed terms in expression (27) are more likely to be positive, and vice versa. The convexity of $B(\pi)$ can arise from two general sources. First, "larger" inventions may be more likely to generate or be more intensive in R&D spillovers.[10] Modeling this important extension is beyond the scope of the paper, however. Second, the demand elasticity for products derived from larger inventions may be lower than for more marginal inventions. In the absence of full appropriation by the inventor, this can generate a convex welfare function. For example, in the pharmaceutical industry it may be argued that a few new drugs targeted at large markets (new cures) are socially more valuable than many smaller improvements on existing drugs that generate the same total private returns to the firm. This implies a convex $B(\pi)$ and makes it more likely that the government

[9] Guesnerie and Laffont (1984) show how the optimal mechanism is obtained.

[10] The empirical literature documents R&D spillovers (e.g. Jaffe, 1986; Bernstein and Nadiri, 1989; Jaffe, Henderson and Trajtenberg, 1993), but there is no yet any evidence on whether the wedge between social and private returns is positively related to the level of private returns.

should set differentiated patent lengths. By allowing longer patents and hence more than proportional increases in profits for larger inventions, such a policy induces firms to tilt their R&D activity toward producing such inventions.

The other key determinant is how the deadweight loss varies with the size of the invention. This is closely linked to the pricing behaviour of the patentees, which depends partly on government restrictions on patent licensing and other factors affecting the appropriability environment. These issues are explored in Section 8.

6. Implementation of the Optimal Mechanism by Simulation Analysis

We have emphasised that the optimal patent mechanism proposed in this paper can be a practical policy tool. Implementation of the mechanism requires that the government know only three key components: the welfare function, the deadweight loss function, and the distribution of the R&D productivity parameter. Moreover, it is important to recognise that setting the optimal *uniform* patent length would require this same information, so that the optimal patent mechanism imposes no additional informational requirements. We believe that it is possible in practice to obtain (at least approximations) to this information and thus to implement the optimal mechanism.[11] This

[11] We recognise that the welfare and deadweight loss functions, and the distribution of θ, may vary across industry groups (reflecting differences in technology, demand and appropriability). Thus the government might wish to tailor the specifics of the mechanism (i.e., the optimal patent length and fee schedules) across industries, if the necessary information were available and the government could enforce such differences.

section uses simulation analysis to illustrate how the optimal patent mechanism can easily be implemented. The analysis is conducted on the model, using various parameterisations of the welfare and deadweight loss functions and the distribution of θ.[12] We compare the optimal patent and fee schedules from these simulations with existing statutory patent lengths and renewal fees in France, Germany and the United Kingdom. In addition, we compute the welfare gains generated by the optimal patent mechanism with differentiated patent lengths, compared with an optimal uniform patent system (with no renewal fees).

The welfare and deadweight loss functions are specified as $W(\pi) = \pi^\alpha$ and $D(\pi) = \beta\pi^\gamma$. Experimentation indicated that incentive compatibility and second order conditions are more easily satisfied when $\alpha > 1$, $\alpha \geq \gamma$ and β is sufficiently large (depending on α). Thus, the simulations were run for a wide variety of (α, β, γ) parameters that satisfy these conditions.[13] The simulation procedure is as follows. For each value of the R&D parameter, θ, the optimal patent length $T(\theta)$ is solved from the first order condition (25), and the second order and incentive compatibility conditions are checked. The optimal fee schedule $f(\theta)$ is computed from equation (24). The total welfare generated by the invention, $W(T, \pi)$, is then computed from equation (21). Fi-

[12] A more complete simulation and policy analysis, incorporating the learning and appropriability issues discussed in Section 7 and 8, will be developed in another paper.

[13] We examined in detail the range of parameters $\alpha \in (1.0, 2.0)$, $\beta \in (2, 10)$ and $\gamma \in (0.5, 2.0)$. Combinations of parameters satisfying incentive compatibility and second order conditions were identified by experimentation. This range of parameters encompasses diverse economic characteristics, including the degree of convexity of the welfare function (α), the ratio of the deadweight loss to profits (β and γ), and the way in which this ratio varies with the size of the invention (γ).

nally, we calibrate the distribution of θ to be broadly consistent with the observed distribution in American manufacturing industries. Recall that θ is the ratio of profits to R&D, $\theta = \frac{\pi}{z}$. The distribution of θ is assumed to be skewed to the right, reflecting a tail of highly R&D productive firms, with a mean value of 3.75. This mean is very close to the (weighted) average ratio of cash flow to R&D in US manufacturing for the period 1983-87 (based on Compustat data), which is 3.7.[14] The mean value corresponds to an average ratio of R&D to profits of about 25%.

Figures 1-3 summarise the results for selected parameter values, but we emphasise that the key findings reported here are robust to variations within the parameter range examined. For each set of parameters we present: (1) the optimal patent schedule $T(\theta)$, (2) the optimal patent fees $f(T)$, and (3) the ratio of $f(T)$ to the optimised present value of profits for each patent length, which we denote $t(T)$. In addition, we report the optimal uniform patent length, T_U, and the percentage improvement in welfare (aggregated over all values of θ) from introducing the optimal patent mechanism, denoted by ΔW.

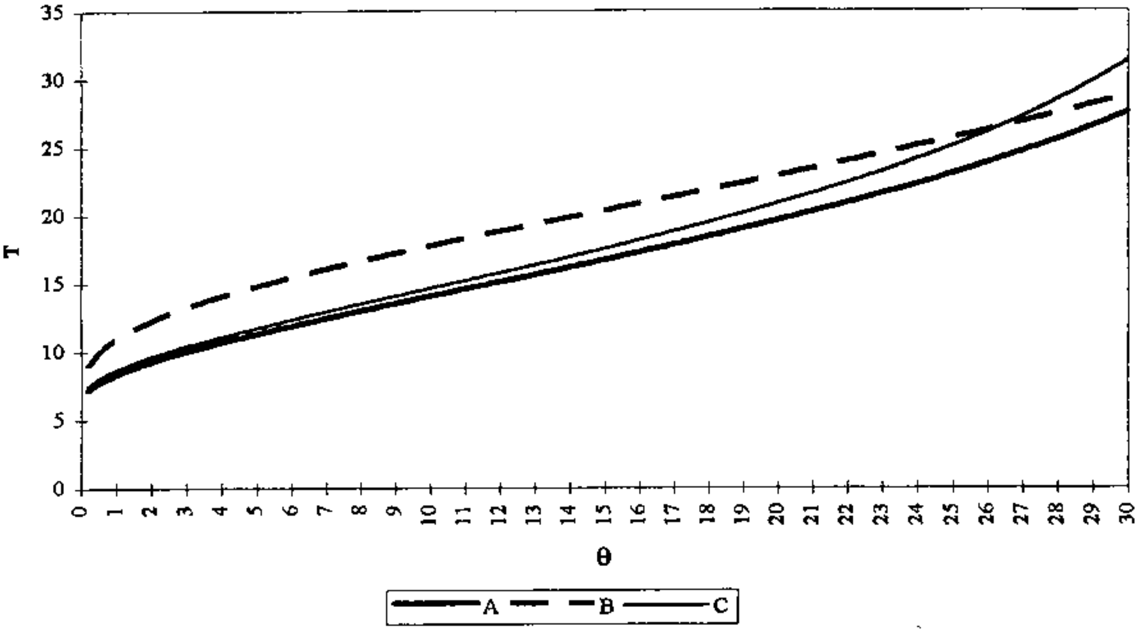
Note first that the simulations indicate an optimal *uniform* patent length of between 15 and 19 years, which is similar to the statutory life-spans in most countries. This provides some support for the plausibility of the parameterisations used here. Turning to the optimal

[14] Two points should be noted. First, cash flow is defined here as operating income plus depreciation minus taxes. For details see Hall (1992). Second, we represent the distribution of θ by a series of five uniform distributions over the range $\theta \in (0,30)$. Specifically, we put 20% of the mass between 0.2 and 2, 55% between 2 and 4, 20% between 4 and 6, and the remaining 5% between 6 and 30. The simulation are conducted over the grid of θ at intervals of 0.2.

differentiated patent schedules, the simulations identify three striking features which are not sensitive to the choice of parameters. First, there is a minimum patent length (about 7 years), even for very low values of θ. This reflects the fact that, while the social value is low for such small patents, so too is the deadweight loss and the R&D inducement effect justifies the patent protection. Second, for the bulk of the distribution of θ, the range of optimal patent lives is actually quite narrow, typically between 8 and 15 years. However, the third feature is that optimal patent lives are much longer for very high values of θ. All existing patent systems impose a maximum life of 20 years or less. While the simulations suggest that there may be relatively few patents which warrant longer optimal patent lives, these are the patents with the greatest contribution to welfare. *Aggregate* welfare is raised by about 2 to 6% by the introduction of the optimal patent mechanism, compared to the optimal uniform patent length. Moreover, the welfare gain rises sharply with the convexity of the welfare function, α, and with the rate at which the welfare to deadweight loss ratio declines with profits, $\alpha - \gamma$, and was in excess of 10% for some parameter values examined.

The optimal (lump-sum) patent fees rise sharply with the patent length. The gradient of Figure 2 corresponds to the annual *renewal* fee, which is seen to rise rapidly for patent lengths up to about 20. This feature is qualitatively consistent with existing statutory renewal fees schedules (see Schankerman and Pakes, 1986). However, the optimal patent fees rise more rapidly than the associated profits from the patent, with the consequence that the optimal "tax" on profits from

Figure 1



	α	β	γ	Tu	%ΔW
A	1.1	5	0.9	15	5.8
B	1.1	2	1	19	2.4
C	1.2	7	1	17	7.0

Figure 2

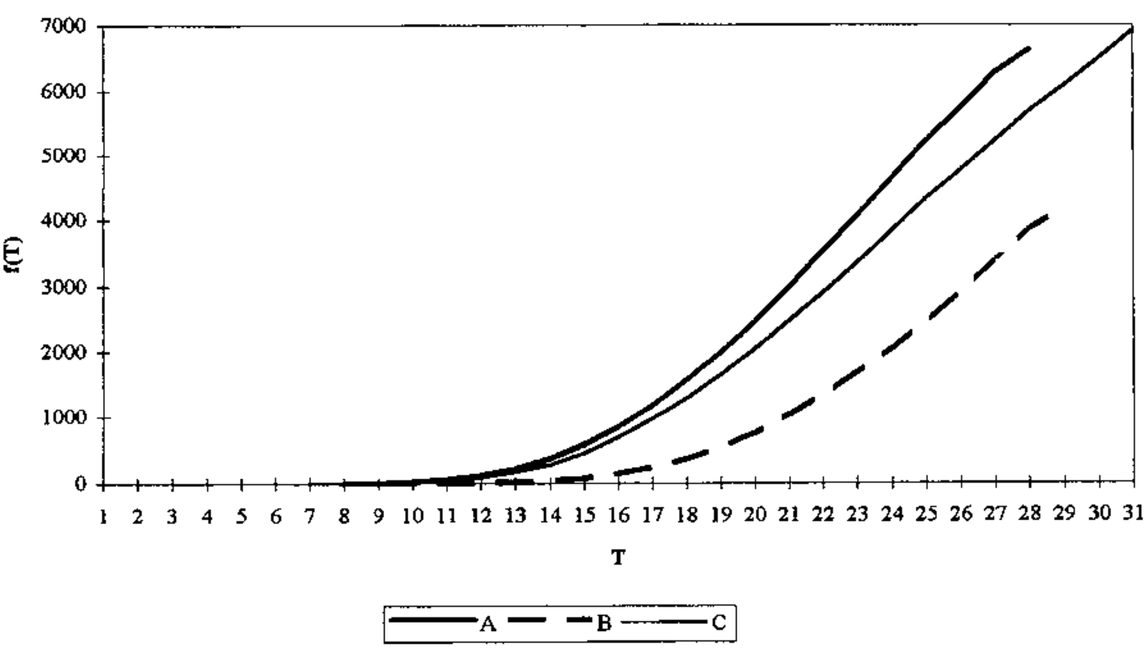


Figure 3

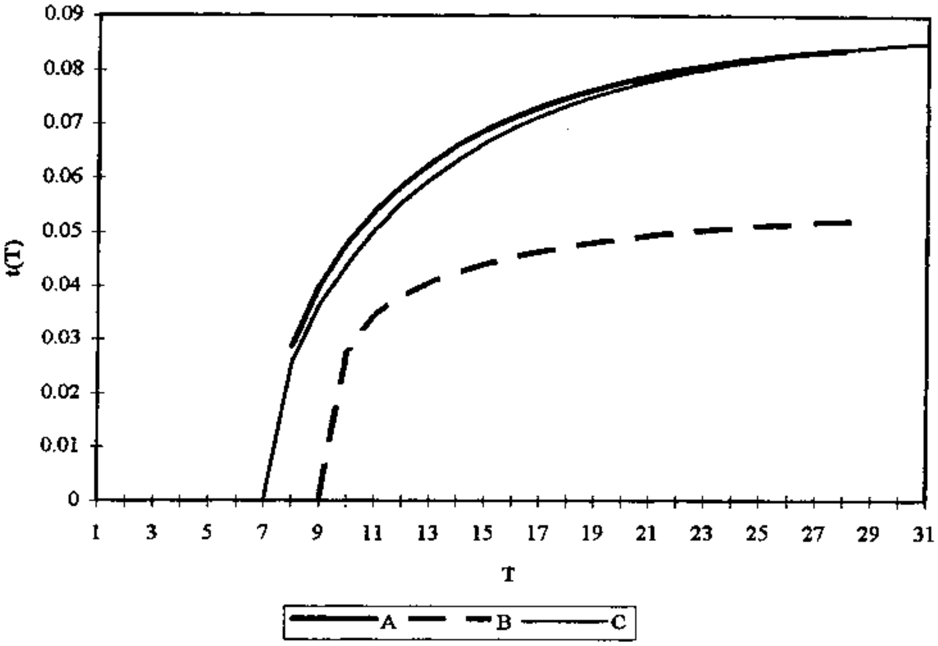
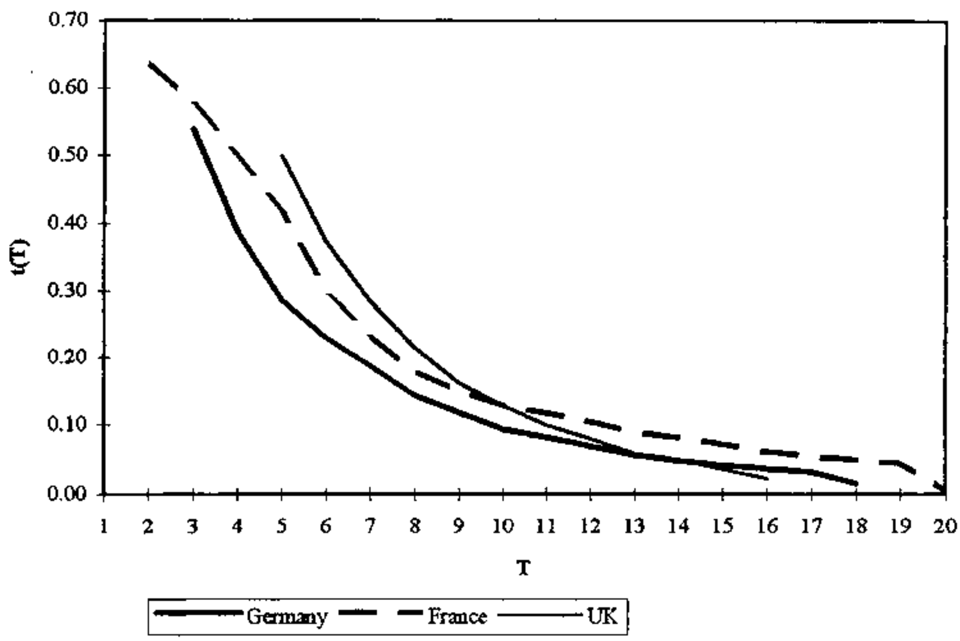


Figure 4



patents, $t(T)$, is sharply progressive, as shown by Figure 3. This important feature of the optimal patent mechanism is sharply violated by existing renewal fees schedules. In order to make the comparison, we use the estimates of the value of patent rights from Schankerman and Pakes (1986) to derive the ratio between actual cumulative renewal fees and the associated profits from patent rights for France, Germany and the United Kingdom.[15] Figure 4 shows that, for all three countries, the actual "tax" on profits from patents is sharply regressive, declining from about 50% for patents canceled at early ages to less than 1% for those kept until the statutory limit.

In short, two central conclusions emerge from this simulation analysis. First, optimal patent lengths should extend beyond the typical statutory maximum values. Second, optimal renewal fees should rise much more sharply with patent length than the existing statutory fee schedules.

7. Post-Patent Learning

In the previous sections we assumed that firms have perfect information on the profits they can obtain from their innovation at the time

[15] Following the procedure in Schankerman and Pakes (1986), the parameter estimates from their patent renewal model (column 2 in Table 3, for each country) are used to characterise the distribution of initial returns to patent protection. We take 500 random draws from this distribution and, using the estimated depreciation rate and the observed renewal fees, compute for each draw the optimal cancellation date (or the statutory maximum, whichever is earlier). We then compute the ratio between the present value of the renewal fees and the returns from patent protection until that cancellation date. The weighted average of this ratio for all patent renewed to each given patent length is reported in Figure 4. Note that renewal fees are required from patent ages 2-20 in France, 3-18 in Germany and 5-16 in UK.

when they apply for the patent. In other words, there is no additional uncertainty once the patent is obtained. However, in reality firms do not have perfect information about the value of their invention: they only have a prior distribution on their returns which they update as they learn during the early life of the patent. Econometric studies document such post-patent learning and indicate that learning is largely completed within four or five years (Pakes 1986, Lanjouw 1992). In this section we introduce this element in a simple way and show, first, that in such case a patent renewal scheme is superior to one in which the firm has to choose patent length and pay a lump sum fee up front , and second, that the optimal renewal fee schedule will depend on the stochastic process governing *ex post* profits.[16]

To do this we modify the timing of the model in the following way. In period 1 firms apply for the patent, but they do not know exactly the profits per period that they will be able to obtain with the patent; π therefore represents only the *expected* profits per period. To simplify the analysis, we assume that after a period of length τ firms learn with certainty the value of the profits per period, which we denote by π_2.[17] We also assume that π_2 is correlated with π and is distributed according to the conditional distribution function $G_2(\pi_2 \mid \pi)$, with density $g_2(\pi_2 \mid \pi)$, in the interval $[0, \bar{\pi}]$. A high π makes a high π_2 more likely, in the sense of first-order stochastic dominance: $\frac{\partial G_2}{\partial \pi} \leq 0$.

[16] The setup in this section also covers the case of "obsolescence", where there is some (exogenous) probability each period that the firm's revenues from the invention fall to zero. Lanjouw (1992) provides the first econometric estimates of obsolescence using patent renewal data.

[17] For simplicity, we will call the period starting after τ period 2. The analysis could be extended to allow for continuous learning in a multiperiod framework, at the cost of considerable complexity. The basic conclusions in this section would remain unchanged.

The Revelation Principle holds also in this setting (see Townsend, 1982) and therefore we can focus on the direct mechanism. The government sets up at the beginning of period 0 a mechanism in which each agent makes announcements $\hat{\theta}$ in period 1—and therefore the related π—and $\hat{\pi}_2$ in period 2. The mechanism specifies a fee $f_1(\hat{\theta})$ to be paid in order to have a patent for the length $T_1(\hat{\theta})$, and in the second period lengths $T_2(\hat{\theta}, \hat{\pi}_2)$ and fees $f_2(\hat{\theta}, \hat{\pi}_2)$. For the sake of simplicity, we restrict the mechanism so that $T_1(\hat{\theta}) \leq \tau$. In other words, the firm can either choose a length shorter than τ—in which case the issue of renewal never arises—or it has to choose whether to renew or not at date τ.[18]

A firm therefore will choose z in order to maximize its net profits. A firm will first choose whether the length will be longer or shorter than τ. Let us focus on the case in which the firm chooses to arrive at least to τ (the other case will be identical to what has already been done in the previous sections):

$$\begin{aligned} \max_z \quad & \int_0^\tau \theta z e^{-rt} dt - f_1(\theta) - \frac{1}{2} z^2 + \\ & + \int_0^{\bar{\pi}} \left[\int_\tau^{T_2(\theta, \pi_2)} \left(\pi_2 e^{-rt} dt - f_2(\theta, \pi_2) \right) g(\pi_2 \mid \pi) \right] d\pi_2 \end{aligned} \tag{28}$$

The first order condition is

$$\frac{\theta}{r} \left[1 - e^{-r\tau} + \int_0^{\bar{\pi}} (\pi_2 - f_2(\theta, \pi_2)) \frac{\partial g}{\partial \pi} (e^{-r\tau} - e^{-rT_2(\hat{\theta}, \hat{\pi}_2)} d\pi_2 \right] - z = 0 \tag{29}$$

which defines the optimal $z^*(\theta, T_2, \frac{\partial g}{\partial \pi})$. The difference with the previ-

[18] In a more general model, where firms learn over time, the issue of the timing of the renewal arises. However, this issue is beyond the scope of the paper.

ous sections is that the optimal R&D effort also depends on how much the *ex post* level of profits is correlated with the *ex ante* (expected) profit which depends directly on z. Therefore, the optimal schedule will depend on $\frac{\partial g}{\partial \pi}$, i.e. the stochastic process generating returns to invention.

Since the purpose of this section is simply to show that the renewal system is superior and to give an example of how the optimal schedule modifies, we do not develop this general framework, but illustrate the line of argument by focusing instead on a special case.[19] Let us assume that profits in the second period take two possible values, zero with probability $\frac{1}{2}$ or 2π with probability $\frac{1}{2}$. It is also quite natural to assume that $w(0) = B(0) = D(0) = 0$. Then we can rewrite (28) as

$$\begin{aligned} \max_z \int_0^\tau \theta z e^{-rt} dt - f_1(\theta) \; & - \; \frac{1}{2} f_2(\theta, 0) - \frac{1}{2} z^2 + \\ & + \; \frac{1}{2}\left[\int_\tau^{T_2(\theta,2\pi)} 2\pi e^{-rt} dt - f_2(\theta, 2\pi)\right] \end{aligned}$$

$$= \frac{\theta}{r} z(1 - e^{-rT_2(\theta,2\pi)}) - f_1(\theta) - \frac{1}{2}\left[f_2(\theta,0) + f_2(\theta,2\pi)\right] - \frac{1}{2} z^2$$

from which we obtain

$$z^* = \frac{\theta}{r}(1 - e^{-rT_2(\theta,2\pi)}) \tag{30}$$

---

[19] A general solution, for a different application, can be found in Laffont and Tirole, 1994.

The welfare maximization problem becomes

$$\max_{T,f_1,f_2} \int_0^{\bar{\theta}} \left\{ \frac{1}{2} \left[\frac{w(2\theta z^*(\theta)}{r} + \frac{D(2\theta z^*(\theta))}{r} e^{-rT_2(\theta,2\pi)} - z^*(\theta) \right] \right\} dF(\theta) \tag{31}$$

subject to the individual rationality constraint of the second period

$$\frac{\pi_2}{r} \left[e^{-r\tau} - e^{-rT_2(\theta,\pi_2)} \right] - f_2(\theta,\pi_2) \geq 0 \tag{32}$$

and of the first period

$$\begin{aligned} EU(\theta,\hat{\theta}) &= \frac{\theta^2}{2r^2}(1 - e^{-rT(\hat{\theta},2\pi)})^2 - f_1(\hat{\theta}) + \\ &- \frac{1}{2}\left[f_2(\hat{\theta},0) + f_2(\hat{\theta},2\pi) \right] \geq 0 \end{aligned} \tag{33}$$

and the truthtelling constraints. The second period incentive compatibility constraint requires that the firm has no incentive to misrepresent π_2 given its report of $\hat{\theta}$.

$$f_2(\hat{\theta},2\pi) \geq f_2(\hat{\theta},0) \ \ \forall \hat{\theta} \tag{34}$$

and

$$\frac{2\pi}{r} \left[e^{-rT_2(\hat{\theta},2\pi)} - e^{-rT_2(\hat{\theta},0)} \right] \geq f_2(\hat{\theta},2\pi) - f_2(\hat{\theta},0) \ \ \forall \hat{\theta} \tag{35}$$

The first period incentive compatibility requires that the firm has no incentive to misrepresent θ given that it anticipates to report truthfully in period 2.

$$\theta \equiv \operatorname*{argmax}_{\hat{\theta}} EU(\theta,\hat{\theta}) \tag{36}$$

The individual rationality constraint of the second period for $\pi_2 = 0$ implies that $f_2(\hat{\theta},0) = 0$. Moreover, $T(\hat{\theta},0)$ appears only in the incentive compatibility constraint in equation (35), and it is clear from there that it is optimal to set it equal to τ (i.e. to renew the patent for zero periods), since it relaxes the constraint and has no effect on the welfare function. Then we are left to determine $f_1(\theta)$, $f_2(\theta, 2\pi)$ and $T_2(\theta, 2\pi)$. Before doing so, however, note that since the optimal direct mechanism involves setting $f_2(\hat{\theta}, 0) = 0$, the implementation of this mechanism requires that the government leaves firms the option to abandon their patent in case the profits turn out to be low. In other words, it is optimal for the government to use a patent renewal scheme rather than an *ex ante* payment scheme.[20]

We next characterize the associated renewal fee. From both equations (32) and (35) we can obtain that

$$f_2(\hat{\theta}, 2\pi) = \frac{2\theta^2}{r^2}\left(e^{-r\tau} - e^{-rT_2(\hat{\theta},2\pi)}\right)\left(1 - e^{-rT_2(\hat{\theta},2\pi)}\right) \tag{37}$$

We want to see how post-patent learning changes the optimal fee schedule . We focus on firms with values of θ such that in the first period they will all choose a length of at least τ.[21] Then for these firms $\frac{df_1}{d\theta} = 0$ and from the maximization in equation (36) we obtain

$$\frac{df_2}{dT_2} = 2\frac{\theta^2}{r}\left(1 - e^{-rT_2(\theta,2\pi)}\right)e^{-rT_2(\theta,2\pi)} \tag{38}$$

[20] One interesting theoretical extension is to allow firms to have different information about the value of their invention at the patenting date. Firms that have relatively more initial information, or learn more quickly, would prefer to trade off large payments in earlier periods for lower renewal fees later. Thus the optimal mechanism might involve offering a *menu* of renewal fee schedules from which firms choose.

[21] If they declare a $\hat{\theta}$ such that $T_1(\hat{\theta}) < \tau$ then nothing changes with respect to the case with no post-patent learning.

Compared with the case without learning, the optimal renewal fees rise twice as quickly with the patent length.[22] In more general setups, the optimal fee schedule will also be more sharply graduated when there is post-patent learning, compared to the non-learning case. This scheme provides the firm with the option not to renew when the realization of profits is low, but in exchange patent fees rise more sharply for those who choose to renew and allow the government to extract part of the additional information rent that accrues through learning.

8. Appropriability Environment

The analysis thus far has been conducted for a given appropriability environment. This section examines how variations in appropriability affect optimal patent policy. Appropriability is heavily influenced by antitrust policy, especially by the treatment of patent licensing (see Gallini and Trebilcock, 1995). Limitations on the licensing arrangements used by patentees reduce the R&D incentive provided by patents but may, if properly designed, increase the use of the invention and hence reduce *ex post* deadweight loss. Whether this occurs depends on the licensing restriction (for discussion see Gallini and Trebilcock, 1995). In the extreme case, a perfectly price discriminating patentee would create no deadweight loss at all, and license restrictions will create such losses as well as blunt R&D incentives. But in other cases, such as refusal by a patentee to license, restrictions can

[22] From the maximization in equation (24)), we obtain for the no-learning case, $\frac{df}{dT} = \frac{\theta^2}{r}(1 - e^{-rT(\theta)})e^{-rT(\theta)}$.

yield offsetting static efficiency gains. It is therefore not clear *a priori* how such limitations, and the appropriability environment more generally, would affect the design of optimal patent life-spans.

The degree of appropriability is defined here as a parameter λ which affects the fraction of the social benefits generated by the invention which is received by the firm: $\frac{\pi}{B}$, where $0 \leq \lambda \leq 1$. In the first approach to this question we assume that λ is common to all firms and, for simplicity, that it does not vary with the size of the invention. Firm-specific differences in appropriability are already captured in the R&D productivity parameter, θ.

The model requires two modifications, the specification of the R&D process and the welfare and deadweight loss functions. We modify the R&D process as follows: $\pi = \lambda\theta z$. The appropriability parameter operates like a "tax" on the returns to R&D. Note that this multiplicative specification implies that greater appropriability increases the incentive to do R&D more strongly for higher productivity firms.[23]

The deadweight loss and the social benefits during the patent life will depend both on the size of the invention (as before) and the appropriability parameter. The total benefits $B(\mu)$ do not depend on the degree of appropriability λ, given the size of the invention. This implies that as λ changes, the function W and D should change in the opposite directions. For a given invention, greater appropriation increases welfare and reduces deadweight loss, except in the unlikely

[23] Following the earlier analysis, the firm chooses R&D level $z^* = \frac{\lambda\theta}{r}(1 - e^{-rT})$, and the present value of maximized profits net of R&D costs are $\pi^* = \frac{(\lambda\theta)^2}{2r^2}(1 - e^{-rT})$. Hence, $\frac{\partial \pi^*}{\partial \lambda}$ is increasing in θ.

case where the increase in rents to the firm comes only from infra-marginal users of the invention (in which case the deadweight loss does not decrease). We therefore write the functions as $\hat{W}(\mu, \lambda)$ and $\hat{D}(\mu, \lambda)$, with $\frac{\partial \hat{W}}{\partial \lambda} > 0$ and $\frac{\partial \hat{D}}{\partial \lambda} < 0$. As in the previous sections, we can invert the function $\pi(\lambda, \mu)$, so that the reduced form welfare and deadweight loss function can be written as $W(\pi, \lambda)$ and $D(\pi, \lambda)$. Notice that the signs of $\frac{dW}{d\lambda} = \frac{\partial W}{\partial \lambda} + \frac{\partial W}{\partial \pi}\frac{d\pi}{d\lambda}$ and $\frac{dD}{d\lambda}$ are ambiguous and depend on $\frac{\partial W}{\partial \pi}$ and $\frac{\partial D}{\partial \pi}$.

Following the same analysis as in Section 5, we obtain the first order condition for the socially optimal patent schedule

$$\begin{aligned} R(T, \theta, \lambda) &\equiv \frac{(\lambda\theta)^2}{r}\left[\frac{\partial w}{\partial \pi} + e^{-rT}\frac{\partial D}{\partial \pi} - (1 - e^{-rT(\theta)})\right] - D(T(\theta)) \\ &= 0 \end{aligned} \tag{39}$$

which is a minor modification of equation (25). By the implicit function theorem, the sign of $\frac{dT}{d\lambda}$ is the same as the sign of $\frac{dR}{d\lambda}$. Using equation (26) (with θ replaced by $\lambda\theta$), this can be expressed as

$$\frac{dR}{d\lambda} = \frac{\partial R}{\partial \theta} + \frac{\lambda}{2}\left\{\left[\frac{\partial^2 W}{\partial \pi \partial \lambda} + e^{-rT}\frac{\partial^2 D}{\partial \pi \partial \lambda}\right] - \frac{r}{(\lambda\theta)^2}\frac{\partial D}{\partial \lambda}\right\} \tag{40}$$

We want to know how the optimal patent schedule changes if the government chooses to screen among different firms. Therefore, we assume that $\frac{\partial R}{\partial \theta} > 0$. Since $\frac{\partial D}{\partial \lambda} < 0$ we conclude that $\frac{dT}{d\lambda} > 0$ unless the expression in square brackets is strongly negative. This sign depends on how the marginal impact of appropriability on welfare and deadweight loss varies with the size of the invention. If this interaction is either non-negative, or "small", then a stronger appropriability environment *increases* the socially optimal length of patents.

This result may appear counterintuitive. From an *ex post* perspective (where a specified level of compensation must be paid to the firm), one would expect stronger appropriability to imply shorter optimal patent lives. However, the analysis shows that the incentive aspects of appropriability cut the other way, and can well dominate in this framework.

9. Conclusions

In this paper we have shown how the government can modify its use of patent renewal fees in order to provide incentives to R&D more efficiently. Our approach emphasizes how different types of uncertainty and of heterogeneity among firms are crucial in determining the optimal use of patent fees. We show that the derivation of the optimal patent fees (and differentiation in patent length which they induce) requires an explicit specification of the R&D process and information structure. We also illustrate with simulation analysis how such information can be used to implement the optimal patent mechanism.

The basic findings in this paper are relevant beyond the particular focus and specification in this paper. The general methodology used in this paper, in particular the application of optimal regulation under various forms of asymmetric information, can be extended to policy design in many other contexts (e.g., see Laffont and Tirole, 1994, for an application to pollution regulation). We intend to explore this in other papers.

The model in this paper has been deliberately simplified to bring out the key intuition. The basic reasoning we use to show how socially optimal differentiation can be derived and implemented is robust to the specification, even if the particular indirect mechanism discussed in this paper would be modified. Several important extensions in this line of research are worth noting. First, post-patent learning about the profitability of inventions is an important feature, as documented in the empirical literature. In a simplified setting, we have shown how the optimal mechanism can be modified to incorporate that consideration without affecting the central results in this paper. More detailed study of this issue is warranted. The simulation analysis developed in Section 6 can be extended to study how the parameters of post-patent learning, and uncertainty in the R&D process, affect the optimal patent mechanism. Second, we emphasised in the paper that appropriability conditions are an important determinant of the incentives to do R&D and hence of the optimal mechanism. These conditions are in part idiosyncratic to the firm but also depend critically on the institutional environment in which the firm operates, most importantly antitrust regulations and competition policy more generally. These other policies can directly influence the appropriability of returns from invention and hence indirectly the optimal mechanism. This interaction was discussed briefly in the paper, but we intend to model it more explicitly and explore its implications for coordinated policy-making in future research.

Appendix

Derivation of equation (24)

Define

$$U(\theta) \equiv \max_{\hat{\theta}} U(\theta, \hat{\theta})$$

By the envelope theorem

$$\frac{dU}{d\theta} = \frac{\theta}{r^2}(1 - e^{-rT(\theta)})^2$$

Reintegrating this equation we obtain

$$U(\theta) = \int_0^\theta \frac{s}{r^2}(1 - e^{-rT(s)})^2 ds + K$$

where K is a constant of integration. Using the individual rationality constraint ($U(0) = 0$) and equating $U(\theta, \hat{\theta})$ (defined in equation (22)), computed at $\hat{\theta} = \theta$, to $U(\theta)$ obtained above, we obtain equation (24) in the text.

The second order condition of the maximization in (23) are

$$U_{\theta\hat{\theta}} = \frac{2\theta}{r}(1 - e^{-rT(\theta)})e^{-rT(\theta)}\frac{dT}{d\theta} \geq 0$$

and can be rewritten as

$$\frac{\theta^2}{r}\left[\frac{\partial^2 w}{\partial \pi^2} + e^{-rT}\frac{\partial^2 D}{\partial \pi^2}\right] - 2\frac{\partial D}{\partial \pi} - 1 \leq 0$$

This condition implies that $\frac{dT}{d\theta} \geq 0$.

References

Bernstein, J. and M. Nadiri (1989), "Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality", *Review of Economic Studies*, **56**, pp. 249-267.

Gallini, N. (1992), "Patent Policy and Costly Imitation", *Rand Journal of Economics*, **23**, pp.52-63.

Gallini, N. and Trebilcock, M. (1995), "Intellectual Property Rights and Competition Policy", mimeo, University of Toronto.

Gilbert, R. and Shapiro, C. (1990), "Optimal Patent Length and Breadth", *Rand Journal of Economics*, **21**, pp. 106-112.

Guesnerie, R. and J.-J. Laffont (1984), "A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm", *Journal of Public Economics*, v. 25, pp.329-369.

Hall, B. (1992), "Investment and R&D at the firm level: Does the source of finance matter?" *NBER working paper*, June, No.4096.

Jaffe, A. (1986), "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value", *American Economic Review*, **76**, pp. 984-1001.

Jaffe, A., M. Trajtenberg and R. Henderson (1993), "Geographical Localization of Knowledge Spillovers as evidenced by patent citations", *Quarterly Journal of Economics*, **108**, pp. 577-598.

Klemperer, P. (1990), "How Broad Should the Scope of a Patent Protection Be?", *Rand Journal of Economics*, **21**, pp. 113-130.

Laffont, J.J. and J. Tirole (1993), "A Theory of Incentives in Procurement and Regulation", Cambridge: MIT Press.

Laffont, J.J. and J. Tirole (1994), "Pollution Permits and Compliance Strategies", mimeo, Toulouse.

Lanjouw, J. (1992), "The Private Value of Patent Rights, a Dinamic Programming and Game Theoretic Analysis of West Germany Patent Renewal Data, 1953-88" Ph.D. Dissertation London School of Economics.

Myerson, R. (1981), "Optimal Auction Design", *Mathematics of Operations Research*, **6**, n.1, pp. 619-632.

Nordhaus, W. (1969), "Invention, Growth and Welfare", Cambridge: MIT Press.

Pakes, A. (1986), "Patents as options: Some estimates of the value of holding European patent stocks", *Econometrica*, **54**, pp. 755-784.

Schankerman, M. (1991), "How valuable is patent protection" *NBER working paper*, July.

Schankerman, M. and A. Pakes (1986), "Estimates of the value of patent rights in European countries during the post-1950 period", *Economic Journal*, **96**, pp. 1052-1077.

Townsend, R. (1982), "Optimal multiperiod contracts and the gain from enduring relationships under private information", *Journal of Political Economy*, **90**, pp. 1166-1186.

Wright, B. (1983), "The Economics of Invention Incentives: Patents, Prizes and Research Contracts", *American Economic Review*, **73**, pp. 691-707.