Patent renewals and R&D incentives*

Francesca Cornelli
(London Business School and CEPR)

Mark Schankerman
(London School of Economics, EBRD and CEPR)

Revised September 1998

Abstract. In a model with moral hazard and asymmetric information, we show that it can be welfare-improving to differentiate patent lives when firms have different R&D productivities. A uniform patent life provides excessive R&D incentive to low-productivity firms, and too little to high ones. The optimally differentiated patent scheme can be implemented through a menu of patent lives (or renewals) and associated fees. We characterise the optimal mechanism, and use simulation analysis to compare it with existing patent renewal systems and to illustrate the potential welfare gains from the optimal policy.

*We are grateful to two anonymous referees and the Co-editor, Rob Porter, for very constructive comments. We also thank Leonardo Felli, Nancy Gallini, Bronwyn Hall, Martin Hellwig, Haizhou Huang, Paul Klemperer, Jenny Lanjouw, Eric Maskin, Ariel Pakes, Suzanne Scotchmer, Jean Tirole and participants at various seminars for comments and Orli Arav for research assistance. All remaining errors are ours. Francesca Cornelli thanks CREST-LEI, Tel Aviv University and the Center for Economic Studies in Munich for hospitality during part of this work. An earlier version of this paper was circulated under the title "Optimal Patent Renewals" (LSE, June 1996).

1. Introduction

Most patent systems require payment of a series of renewal fees in order to maintain patent protection up to the statutory patent life. Typically more than half of all patents are voluntarily cancelled by nonpayment within ten years of the date of patent application. Thus, even though all countries impose a uniform statutory patent life, there is *de facto* differentiation in patent lives. Econometric studies have confirmed that renewal fees influence the decision to patent and that more valuable patents are held longer (Pakes 1986, Schankerman and Pakes 1986, Schankerman 1998 and Lanjouw 1998). However, in practice patent renewal fees are used to finance patent offices and there is no reason to believe that the existing pattern of *de facto* patent lives induced by these fees improves welfare.

The central idea in this paper is that patent fees can be used as an incentive device to implement a policy of *optimally* differentiated patent lives (and more generally, differentiated patent protection). To demonstrate the potential benefits of using patent renewal fees in this way, we show how differentiated patent lives can be better, in terms of social welfare, than a uniform patent life.

The use of patents as a policy instrument to provide R&D incentives makes sense only if there is private information about the cost or value of inventions (Wright 1983). We develop a static model of innovation that incorporates both asymmetric information on cost (R&D productivity) and moral hazard on the R&D effort undertaken by the firm. The basic intuition in the paper is that differentiated patent lives can be welfare-improving because of an "incentive effect": allowing firms with high R&D capabilities to choose longer patent lives gives these firms an incentive to invest more R&D resources. Any uniform patent life will provide too much incentive to low R&D-productivity firms and too little incentive to high ones. This generates both a suboptimal level and distribution of R&D. We believe that this basic point will carry over to a dynamic framework involving sequential innovation, but this is not explored here.

Differentiating patent lives can improve social welfare when there is *ex post* heterogeneity in the value of inventions that the government cannot observe. This can arise from *ex ante* heterogeneity in R&D productivity or from uncertainty in the research process. The optimal scheme involves the government offering firms an incentive-compatible menu of patent lives and associated lump-sum patent fees. Each firm chooses its preferred patent life. This menu of patent lives is equivalent to the government offering a schedule of annual renewal fees, if there is no post-patent learning by the firm. However, we show that if the firm learns about the value of its invention after patenting, the renewal scheme is superior in terms of social welfare.

We use simulations to illustrate what the optimal differentiated patent scheme might look like and to compare it to existing patent systems. We find four striking characteristics: there is a minimum patent life even for small inventions; for most inventions the range of optimal patent lives is quite narrow; optimal patent lives are much longer for particularly good inventions (much greater than existing statutory maximum lives); and patent fees should rise much more sharply with patent life relative to the existing fee schedules.

In contrast to this paper, most of the traditional patent-design literature has focused on the optimal *uniform* patent length and, more recently, other dimensions of patent policy such as breadth (Nordhaus 1969, Klemperer 1990, Gilbert and Shapiro 1990, Green and Scotchmer 1995, Scotchmer 1996, O'Donoghue, Scotchmer and Thisse 1998). There are only two recent papers that study differentiated patent protection in different frameworks. Scotchmer (1997) analyses a static model with private information on the cost and value of inventions, but no moral hazard (the firm chooses which ideas to develop, but not how much R&D to do). She shows that asymmetry of information is sufficient to justify the use of patents to provide R&D incentives, and that any direct mechanism can be implemented using a renewal mechanism. de Laat (1997) analyses a patent race in which the imitation delay is private information, and studies optimal differentiation of patent length and breadth.

The paper is organised as follows. Section 2 presents the model. Section 3 characterises the second-best patent policy and the optimal uniform length, which serve as benchmarks for the welfare analysis. Section 4 presents the main result: the conditions under which it is optimal to differentiate patent lives, and the characterisation of the optimal scheme. In Section 5 we simulate the optimal mechanism and the associated welfare gains, and compare the optimal patent lives and fees to existing patent systems in France, Germany and the United Kingdom. In Section 6 we briefly discuss three extensions of the model: stochastic R&D outcomes, post-patent learning and the appropriability environment. Concluding remarks summarise the key findings and suggest directions for future research.

2. The model

The timing of the model is as follows: at the beginning of the first period, firms decide how much to invest in R&D, which yields an invention at the end of the period. In the second period, firms choose a patent life of length T. The invention continues to be used indefinitely after patent expiration. Let μ denote the 'size' of the invention which affects the level of demand, $Q(p \mid \mu)$, with $Q_\mu > 0$ (subscripts denote partial derivatives). We assume that the innovating firm charges a uniform monopoly price while the patent is in force, and that the competitive price prevails after the patent expires. These assumptions simplify the analysis, but the argument in the paper also holds for more complex appropriability environments, such as allowing for price discrimination and licensing (see Section 6).

Without loss of generality we set marginal cost to zero. The firm with the patent sets price p to maximise flow revenue $pQ(p \mid \mu)$. Let $p^*(\mu) \equiv \text{argmax}_p pQ(p \mid \mu)$ and flow profits during the patent life be $\pi(\mu) \equiv p^*(\mu)Q(p^*(\mu) \mid \mu)$. Since by the envelope theorem $\frac{d\pi}{d\mu} = p^*(\mu)Q_\mu > 0$, $\pi(\mu)$ is monotonic increasing and we can express μ as a function of the level of (maximised) profits: $\mu = v(\pi)$. The appropriability assumptions are embedded in the function $v(\pi)$. More effective appropriation by the firm would shift the $v(\pi)$ function down since it allows a 'smaller' invention to generate the same profit. We refer to π as the 'output' of the R&D, but it should be interpreted as a summary statistic for μ, given the capacity of the firm to appropriate the surplus, which determines the relationship between μ and π. All that matters for the subsequent analysis is that we can express profits as a monotonic function of the size of the invention.[1]

The firm maximises total profit given by

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

Flow welfare (profits plus consumer surplus) during the patent life is

$$W(\pi) \equiv \int_{p^*(v(\pi))}^{\infty} Q(p \mid v(\pi)) dp + \pi \tag{2}$$

After the patent expires, profits are zero and flow welfare is

$$B(\pi) \equiv \int_0^{\infty} Q(p \mid v(\pi)) dp \tag{3}$$

Note that $B(\pi) = W(\pi) + D(\pi)$ where

$$D(\pi) \equiv \int_0^{p^*(v(\pi))} Q(p \mid v(\pi)) dp - \pi \tag{4}$$

is the deadweight loss from the patent. It is easy to show that $D_\pi(\pi)$ cannot be signed, so that 'larger' inventions may generate either larger or smaller deadweight loss. However, we assume that 'larger' inventions generate greater social benefits after the patent expires, so $B_\pi(\pi) > 0$.

The government's objective function is

$$S(\pi, T) = \int_0^{+\infty} W(\pi) e^{-rt} dt + \int_T^{+\infty} D(\pi) e^{-rt} dt = \frac{W(\pi)}{r} + \frac{D(\pi)}{r} e^{-rT} \tag{5}$$

The first term represents profits and consumer surplus for the (infinite) life of the invention; the second term is the additional gain which accrues after the patent expires (formerly deadweight loss).

We turn next to the process generating profits. This involves, first, the production

[1]The analysis also holds for a process invention. In that case, if μ is unit cost savings, π is monotonic in μ, but the level of demand Q is unaffected (given p).

of an innovation and, second, the appropriation of the surplus by the firm in the form of profits. Let z denote R&D input, and the parameter θ reflect both the capacity of the firm to generate an innovation and to appropriate the surplus it generates (hereafter, the "R&D productivity parameter"). The potential surplus from the innovation is indexed by μ, which depends on R&D input and the firm's R&D productivity parameter, i.e. $\mu(\theta, z)$. Profits depend on this surplus and the ability of the firm to appropriate these rents, i.e. $\pi = h(\theta, z)$. Therefore, the marginal productivity of R&D, h_z, is a function of θ: different firms have different marginal productivity in the R&D process. We assume that $h_\theta > 0$, $h_z > 0$ and $h_{zz} < 0$. We also make the natural assumption that the marginal productivity of R&D is nondecreasing in θ: $h_{z\theta} \geq 0$. This ensures that the optimal differentiated patent policy is implementable (see Section 4). To simplify the analysis we assume quadratic costs of R&D effort, $\psi(z) = z^2/2$, but the qualitative results carry over for any non-concave cost function. We will also study the special case $\pi = \theta z$, where firms have a constant marginal productivity θ, which can be interpreted as the ratio of R&D to profit.

The model treats the R&D process as deterministic (the stochastic R&D case is discussed briefly in Section 6). Given its private information on θ, the firm chooses the size of the invention (hence π) by setting R&D input z. The government does not know the value of the invention π (equivalently, θ), but it knows that θ is drawn from the distribution function $G(\theta)$, defined over the interval $[0, \bar{\theta}]$, with density $g(\theta)$.

3. Benchmarks

If the government observed θ and R&D effort, it could enforce the first-best level of R&D effort without resort to patents.[2] This is not achievable because of asymmetric information, thus we study the case where the government uses patents to provide R&D incentives.[3] In this section we characterise the full-information second best

[2]The first-best level of R&D is defined by $z^{FB}(\theta) = \text{argmax}_z\{\frac{B(h(\theta,z))}{r} - \frac{z^2}{2}\}$.

[3]See Scotchmer (1997) for a general argument that asymmetric information is sufficient to justify the use of patents.

patent policy, where the government sets the optimal patent length for each value of θ, and the optimal uniform patent life when θ is private information. These will serve as benchmarks for the subsequent welfare comparisons.

In both cases, the government has to take into account that, given a patent length T, the firm sets R&D to maximize profits. Given R&D effort z and productivity θ, the firm earns profit net of resource costs:

$$\int_0^T h(\theta, z)e^{-rt}dt - \frac{z^2}{2} = \frac{1}{r}h(\theta, z)(1 - e^{-rT}) - \frac{z^2}{2} \tag{6}$$

The first order condition of the maximization of (6) is

$$\frac{1}{r}(1 - e^{-rT})h_z - z = 0$$

Since $h(\theta, z)$ is concave, the second order conditions are satisfied and by the implicit function theorem there exists a function ϕ such that $z^* = \phi(\theta, T)$, with $\phi_T > 0$ and $\text{sign}[\phi_\theta] = \text{sign}[h_{z\theta}]$. Define the optimized profit as $H(\theta, T) \equiv h(\theta, \phi(\theta, T))$ with $H_T > 0$ and $H_\theta \geq 0$ provided $h_{z\theta} \geq 0$. For reference, in the linear case $\pi = \theta z$ we obtain $z^*(\theta, T) = \frac{\theta}{r}(1 - e^{-rT})$.

Proposition 1: *The full-information second-best policy, $T^{**}(\theta)$, satisfies the following equation at each value of θ:*

$$\frac{1}{r}\left[W_\pi + e^{-rT}D_\pi\right]H_T = z^* z_T^* + e^{-rT}D(\pi). \tag{7}$$

Proof: Maximizing $S(\pi, T)$ in (5) subject to the constraint $\pi = H(\theta, T)$ yields the first order conditions in Proposition 1. ■

The policy $T^{**}(\theta)$ equates the marginal social benefit and cost of extending T. The left hand side of (7) is the marginal benefit from extending the patent life T. This reflects both the incentive effect on R&D and thus on the size of the invention (profits), and the marginal social value of larger inventions. The right hand side is the marginal social cost, comprised of the additional R&D cost and the discounted value of the additional period's deadweight loss associated with an increase in T.

Now consider the case where θ is private information, and suppose the government sets the optimal *uniform* patent length. This involves maximizing $\int_0^{\bar{\theta}} S(\pi, T) dG(\theta)$ subject to $\pi = H(\theta, T)$. The optimal uniform length is described in Proposition 2.

PROPOSITION 2: *The optimal uniform patent length, T^U, satisfies the following equation:*

$$\int_0^{\bar{\theta}} \left\{ \frac{1}{r} \left[W_\pi + e^{-rT} D_\pi \right] H_T \right\} dG(\theta) = \int_0^{\bar{\theta}} \left\{ z^* z_T^* + e^{-rT} D(\pi) \right\} dG(\theta). \tag{8}$$

Proof: As in Proposition 1, using the objective function $\int_0^{\bar{\theta}} S(\pi, T) dG(\theta)$. ■

Note that the government must know the distribution of the R&D parameter $G(\theta)$ to set the optimal *uniform* patent length. As Section 4 shows, this information is not needed to set optimal *differentiated* patent lives.

4. Optimal Differentiated Patent Policy

When the firm's R&D productivity is private information, the government may want to provide an incentive structure that shifts the distribution of R&D effort toward the high-θ firms. With the optimal uniform length, the high-θ firm already obtains greater profit from its more valuable invention. But it is not sufficient because, in setting T^U, the government has averaged across θ (see Proposition 2) and thus provides too little incentive to high-θ firms and too much to low-θ firms. The consequence is that the social cost of producing inventions is not minimised. In this section we use a mechanism design approach, derive the optimal patent policy $T^*(\theta)$ and show that under certain conditions it is indeed differentiated. When the optimal patent scheme is differentiated, we characterise two indirect mechanisms to implement this policy. First, the government can offer a menu of patent lengths and associated up-front fees, and firms self-select when they apply for a patent. If the firm does not learn about the value of the invention after it patents, this mechanism is equivalent to a second one, a renewal scheme where the firm makes a sequence of decisions to extend the patent by payment of renewal fees. In Section 6 we introduce *ex post* learning in a simplified version of the model and show that the renewal scheme is superior to a menu with up-front fees.

By the Revelation Principle, we restrict attention to direct mechanisms where the firm announces $\hat{\theta}$ and the government determines the patent length and the fee as a function of the announced value $\hat{\theta}$: $\{T(\hat{\theta}), f(\hat{\theta})\}$.[4] Facing this schedule, the firm chooses R&D effort z to maximize profits in (6), which yields $z^* = \phi(\theta, T(\hat{\theta}))$. Its payoff is therefore given by:

$$U(\theta, \hat{\theta}) = \int_0^{T(\hat{\theta})} h(\theta, z^*)e^{-rt}dt - \frac{1}{2}(z^*)^2 - f(\hat{\theta}) \tag{9}$$

The welfare maximisation problem becomes:

$$\max_{T,f} \int_0^{\bar{\theta}} \left[\frac{W(h(\theta, z^*))}{r} + \frac{D(h(\theta, z^*))}{r} e^{-rT(\theta)} - \frac{1}{2} z^{*2} \right] dG(\theta) \tag{10}$$

subject to

$$\text{IR constraint: } U(\theta, \theta) \geq 0, \ \forall \theta \tag{11}$$

and

$$\text{IC constraint: } \theta = \text{argmax}_{\hat{\theta}} U(\theta, \hat{\theta}), \ \ \forall \theta, \hat{\theta} \tag{12}$$

This is a standard mechanism design problem, except that the fees which implement the optimal policy do not enter the objective function. To solve it, we first use the IR and IC constraints to obtain the fee schedule $f(\theta)$, and then derive $T(\theta)$ from the unconstrained maximization of (10).

The following proposition characterises the optimal patent mechanism:

PROPOSITION 3: *An optimal differentiated patent policy,* $\{f^*(\theta), T^*(\theta)\}$ *has to sat-*

[4] For simplicity, we will assume that T is differentiable. However, the argument can be extended to the case in which T is piecewise differentiable. See Scotchmer (1997) for a discussion in a related set-up.

isfy the following necessary conditions for each θ*:*

(1)

$$R(T,\theta) \equiv \frac{1}{r}\left[W_\pi + e^{-rT}D_\pi\right]H_T - z^* z_T^* - e^{-rT}D(\pi) = 0 \tag{13}$$

(2)

$$f^*(\theta) = H(\theta,T)A(\theta) - \frac{1}{2}(z^*)^2 - \int_0^\theta h_s(s,z)A(s)ds \tag{14}$$

where $z^* = \phi(\theta, T(\theta))$ *and* $A(\theta) \equiv \frac{1}{r}(1 - e^{-rT(\theta)})$.

Proof: See Appendix A1.

In Appendix A1 we derive the first order conditions of the maximization implied by constraint (12) and, following Myerson (1981), obtain the fee schedule that guarantees incentive compatibility, given in equation (14). The first order conditions of the maximization of the objective function in (10) give equation (13). We also derive the sufficient conditions for $\{f^*(\theta), T^*(\theta)\}$ to be optimal—i.e., the second order conditions for the maximization of (10) and the maximization implied by (12). This last condition implies that the optimal patent length schedule is monotonic strictly increasing: $T_\theta^*(\theta) > 0$. Therefore, we must check that the patent policy defined by (14) is increasing. If it is not, the optimal differentiated patent policy is not incentive compatible and the best the government can do is to set T constant.[5]

Remark 1. Since the first order condition in (13) is the same as in Proposition 1, the optimal differentiated patent schedule implements the full information, second-best policy.

Remark 2. Since welfare in (10) is maximised pointwise, the optimal patent schedule $T^*(\theta)$ does not depend on $G(\theta)$. By contrast, the optimal uniform patent length does depend on this distribution.

Remark 3. The patent fees in (14) equal the present value of *maximised* profits net of R&D costs minus the information rent which must be left to the firm to

[5]If (14) implies a non-monotonic schedule $T(\theta)$, T may be constant only for some intervals. See Guesnerie and Laffont (1984).

induce revelation of θ. In Appendix A1 we show that the fees are defined up to a constant. In equation (14) this constant is set equal to zero, so that the fees are as high as possible consistent with the individual rationality constraint for all θ types. Thus, the government is extracting all the inframarginal rent from the firm except the informational rent needed for incentive compatibility. Scaling down all fees by a constant, however, would not change the optimal patent schedule because patent fees are pure transfers that do not affect social welfare in this model. If instead we treated them as a substitute for costly public funds (as in Laffont and Tirole, 1993), optimal patent policy would be closely tied to the shadow price of such funds. It would be straightforward to extend the model in this way. However, in this paper we want to emphasize the R&D incentive aspect of patent policy, since we do not think it is practical to tie patent policy to fiscal conditions.

Note that even when optimal patent fees are the highest level possible consistent with the individual rationality and incentive compatibility constraints, they could be negative for low θ firms. Since incentive compatibility requires $T(\theta)$ to increase in θ, in order to avoid very long patents it may be optimal for the government to set very short lives and subsidize the R&D for low θ firms. But this would involve costly public funds and monitoring costs, to ensure that the subsidised firms actually innovate. These costs would have to be taken into account in the optimal design if patent fees entered the objective function. In that setting, the optimal policy could imply that all the fees are scaled up, to insure they are non-negative for all θs. In this case, firms with low θ values may not undertake any R&D at all. This is equivalent to imposing a minimum standard for patentability.[6]

We now characterise two special cases to highlight the intuition and the key factors shaping the optimal design.

[6]If the R&D process is deterministic, constraining fees to be non-negative implies that some low θ firms will choose not to do R&D since it will not generate sufficient profit to cover the minimum patent fee. However, since the government can only differentiate patent protection on the basis of R&D outcomes, if these are stochastic then some high θ firms may also produce innovations that fall below the threshold. But this will occur less frequently than for low θ firms (in the stochastic dominance sense). There will be second-best, *ex ante* efficiency, but *ex post* inefficiency will arise due to asymmetric information.

COROLLARY 1: *If* $\pi = \theta z$*, then* $z^* = \frac{\theta}{r}(1 - e^{-rT(\theta)})$ *and the optimal differentiated patent policy* $\{T^*(\theta), f^*(\theta)\}$ *satisfies the following necessary and sufficient conditions for each* θ*:*

(1)

$$\frac{\theta^2}{r}\left[W_\pi + e^{-rT}D_\pi - (1 - e^{-rT(\theta)})\right] - D(\pi) = 0$$

(2)

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

(3)

$$-\frac{D(\pi)}{\pi} + D_\pi < \frac{\theta^2}{r}\left[W_{\pi\pi} + e^{-rT}D_{\pi\pi}\right] \leq 1 + 2D_\pi$$

Conditions (1) and (2) are just the conditions given in Proposition 3 for the case $\pi = \theta z$. To derive the inequalities in (3), we substitute the first order condition given in (1) into the two sufficient conditions derived in Appendix A1, point (iii). The LHS guarantees that $T^*_\theta(\theta) > 0$, and the RHS is the second-order condition for the welfare maximization. The inequalities in (3) depend on how welfare and deadweight loss vary with the level of profit. This underlines the fact that optimal patent policy depends on how the distribution of private and social benefits vary with the size of the invention, and thus on any other policies that affect this relationship (such as competition policy constraints on licensing and other forms of appropriation).

If $B(\pi)$ is convex, it is more likely that the left-hand inequality (incentive compatibility) is satisfied. This may arise from two sources. First, "larger" innovations may be more likely to generate R&D spillovers than marginal ones. This remains an unresolved empirical issue.[7] 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

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

example, in the pharmaceutical industry an important new drug targeted at a large market may be *socially* more valuable than many smaller improvements on existing drugs that generate the same *private* returns to the firm. By allowing longer patents and hence more than proportional increases in profits for larger inventions, a differentiated patent policy can induce firms to tilt their R&D activity toward producing such inventions.

COROLLARY 2: *If welfare and deadweight loss are proportional to profits,* $W(\pi) = \alpha\pi$ *and* $D(\pi) = \beta\pi$, *and* $\pi = \theta z$, *then it is optimal to set a uniform patent life equal to*

$$T^U = \frac{1}{r}[ln(1 + 2\beta) - ln(1 + \beta - \alpha)].^8$$

In this special case, $T^*_\theta(\theta) = 0$ and since the second order conditions for welfare maximization are met, the optimal patent policy is constant. However, note that for a generic concave profit function $\pi = h(\theta, z)$ it is not generally optimal to set a uniform patent length even when welfare and deadweight loss are proportional to profits.

We have characterised the optimal direct mechanism. We now show it can be implemented by using two alternative indirect mechanisms, which are observed in practice: an up-front menu of patent lengths and fees, and a renewal scheme. With the up-front menu, a firm chooses among different patent lengths T and pays the corresponding fee $F(T)$. To find the fee schedule that implements the optimal direct mechanism, note that $T^*(\theta)$ is monotone strictly increasing, so we can invert it and obtain $\theta(T^*)$. Substituting this into the fee schedule of the optimal direct mechanism, $f^*(\theta)$, derived in Proposition 3, we obtain the lump sum payment associated with each patent length, $F(T) \equiv f^*(\theta(T))$. The maximum patent life is given by $\bar{T} \equiv T^*(\bar{\theta})$.

To find the renewal mechanism, we need the fee $R(t)$ that a firm must pay in each period t if it wants to renew its patent for that period. The minimum patent

[8] The linear specification requires $\beta > \alpha - 1$, i. e. the marginal deadweight loss must exceed the gain in consumer surplus from an increase in patent life. Otherwise the optimal uniform length is infinite.

life is given by $T^*(0) \geq 0$, which is associated with a fee $R(0) \equiv f^*(0)$. All the other renewal fees for each period $T^*(\theta)$ are such that

$$f^*(\theta) = \int_0^{T^*(\theta)} R(t)e^{-rt}dt.$$

The relationship between the fees of these two indirect mechanisms is given by

$$\frac{dF}{dT} \equiv \frac{df^*(\theta(T))}{dT} = R(T)e^{-rT}.$$

That is, the gradient of the optimal up-front payment schedule is given by the present value of the renewal fee, for each patent length.

To establish equivalence between the direct mechanism and the renewal scheme, we need to show that the incentive compatibility constraint is satisfied and the firm chooses the same level of R&D under the renewal scheme. In the direct mechanism the firm's optimisation and incentive compatibility constraint in (6) and (12) are expressed as functions of θ. Since there is a one-to-one relationship between T and θ, the solution to these maximisation problems will be the same if they are expressed in terms of T, as in the indirect mechanism.[9]

5. Simulating the Optimal Mechanism

In this section we use simulation analysis to illustrate the key features of the optimal differentiated patent mechanism and to compute the welfare gains relative to an 'optimal' uniform patent system. In addition, we compare the patent lives and fees from the optimal mechanism with existing statutory patent lives and fees in France, Germany and the United Kingdom.

To conduct the simulations, we use a linear R&D process $\pi = \theta z$, and specify flow

[9]If the patent schedule $T^*(\theta)$ is discontinuous, this argument requires refinement. Scotchmer (1997) shows that, both in her model and ours, any R&D plan that is implementable with a direct mechanism is implementable with a renewal mechanism, even when $T^*(\theta)$ is discontinuous. This would cover, for example, cases where firms are segmented into two groups (low and high productivity) with different fixed patent lives.

welfare and deadweight loss as $W(\pi) = \pi^{\alpha}$ and $D(\pi) = \beta\pi^{\gamma}$. We run the simulations for a variety of (α, β, γ) values.[10] For each value of θ, we compute the optimal patent length $T(\theta)$ from equation (13) and check that the second order and incentive compatibility conditions are met. The optimal fee schedule $f(\theta)$ is computed from equation (14) and welfare $S(\pi, T)$ from (10). The distribution of θ is calibrated to be broadly consistent with the observed distribution of the ratio of profits to R&D in US manufacturing. The distribution is assumed to be skewed to the right, reflecting a tail of highly productive R&D-performing firms, with a mean value of 3.75. This is very close to the (weighted) average ratio of cash flow to R&D in US manufacturing for the period 1983-87 of 3.7, based on Compustat data.[11]

Figures 1-3 summarise the results for selected parameter values, but the features reported here are robust to parameter variations within the range examined. We present the optimal patent schedule $T(\theta)$, the optimal patent fees $F(T)$, and the ratio of $F(T)$ to maximised profits for each T (the 'equivalent tax rate' on patent returns), which we denote $t(T)$. We also provide the optimal uniform length, T^U, and the percentage gain in welfare from introducing the optimal mechanism, denoted by ΔS.

As Figure 1 shows, the simulations indicate an optimal uniform patent length of between 15 and 19 years, which is very close to the statutory lifespans in most countries.[12] There are three striking and robust features of the optimal differentiated patent schedules in Figure 1. First, there is a minimum length (about 7 years), even for very low values of θ. This reflects the fact that, while the social value for such

[10] Experiments showed that the second order and incentive compatibility conditions are more commonly satisfied when $\alpha > 1$, $\alpha \geq \gamma$ and β is sufficiently large. We examined the parameter space $\alpha \in (1.0, 2.0)$, $\beta \in (2, 10)$ and $\gamma \in (0.5, 2.0)$, which allows for variations in the 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 profits (γ).

[11] 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 $G(\theta)$ by a series of five uniform distributions over the range $\theta \in (0, 30)$: 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. Simulations are conducted over the grid of θ at intervals of 0.2.

[12] Under the recent World Trade Organisation Agreement, signatory countries will harmonise their statutory patent life at 20 years (but not their renewal fees).

small patents is low, so too is the deadweight loss, and the R&D incentive effect justifies the protection. Second, for the bulk of the distribution of θ, the range of optimal patent lives is quite narrow, typically between 8 and 15 years. However, the third feature is that the optimal lives for very high values of θ are much longer than existing statutory limits (20 years or less). And although there may be relatively few patents which warrant longer patent lives, these are the patents with the greatest contribution to welfare.

The optimal mechanism raises aggregate welfare by 2-7%, as compared to the optimal uniform length. The welfare gain depends on two features of the underlying welfare and deadweight loss functions. First, the welfare gain rises with the convexity of the welfare function, α. Second, the gain is larger when the ratio of welfare to deadweight loss declines more quickly with the size of the innovation (profits), i.e. when $\alpha - \gamma$ is higher. The welfare gain from the optimal mechanism exceeds 10% for some parameter values examined.[13]

Figure 2 shows that the optimal (up-front) patent fees rise sharply with patent life. The gradient of the curves corresponds to the annual *renewal* fee, which also rises for patent lengths up to about 20. This feature is qualitatively consistent with existing statutory renewal fee schedules (Schankerman and Pakes, 1986). The optimal patent fees rise more rapidly than the associated profits from the patent, so that the "equivalent tax", $t(T)$, is *progressive*, as shown in Figure 3. This important feature of the optimal mechanism is violated by existing renewal fee 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 profits from patent protection in France, Germany and the United Kingdom.[14] Figure

[13] If α becomes too high, the benefits from larger inventions increase very rapidly and the optimal patent length associated with the highest θs becomes infinite.

[14] We use the Schankerman and Pakes (1986) parameter estimates of 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 value-weighted average of this ratio for all patents renewed to each patent

4 shows that for all three countries renewal fees constitute a *regressive* 'tax' on profits from patents, declining from about 50% for patents cancelled at early ages to less than 1% for those renewed until the statutory limit.

In short, the simulation analysis indicates that optimal patent lives should extend beyond the typical statutory maximum (or the 20-year life required by the World Trade Organization Agreement), and that the optimal renewal fees should rise much more with patent length than existing fee schedules.

6. Extensions

In this Section we briefly introduce three extensions of the model: stochastic R&D outcomes, post-patent learning and appropriability. The purpose is to illustrate how the model can be adapted to incorporate a richer description of the R&D process and post-invention competition. A complete treatment of these issues is left for future work.

6.1. Stochastic R&D Outcomes

In order to introduce a stochastic R&D process, we consider a risk neutral firm with R&D technology $\pi = h(z, \theta) + \epsilon$, where ϵ has distribution $V(\epsilon)$ and is observed only after the firm chooses its R&D. Under the direct mechanism, the firm announces $\hat{\pi}$ and obtains a patent of length $T(\hat{\pi})$. The profit maximising level of R&D, z^*, is given by:

$$\int_\epsilon \left[\frac{h_z(\theta, z^*)}{r} (1 - e^{-rT(h(\theta,z^*)+\epsilon)}) - z^* \right] dV(\epsilon) = 0$$

Then the welfare maximisation problem becomes

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

length is reported in Figure 4. Renewal fees are required for patent ages 2-20 in France, 3-18 in Germany and 5-16 in the United Kingdom.

Given a distribution $V(\epsilon)$, standard simulation techniques can be used to solve this problem. While the characteristics of the optimal patent schedule will depend on $V(\epsilon)$, there will again be conditions under which it is optimal to differentiate patent lives.[15]

6.2. Post-Patent Learning

We have assumed that firms know the value of their innovation when they apply for a patent. However, in reality firms only have a prior distribution on their returns, which they update as they learn during the early life of the patent. Econometric studies have documented such post-patent learning and shown that it is largely completed within four or five years after the patent application date (Pakes 1986, Lanjouw, 1998). We consider a simplified model with post-patent learning and show that, in this case, a patent renewal scheme is welfare-superior to an up-front menu of lengths and fees. We believe that the qualitative result extends to more complex models of learning, but have not formally proved this conjecture.[16]

We modify the timing of the model in the following way. In period 1 firms apply for the patent knowing only the *expected* profit per period, π. After a period of length τ, firms learn with certainty the value of the profit per period, which can take on either the value of zero or 2π with equal probability. We call the period starting after τ period 2, and assume that $\pi = \theta z$.[17]

[15] When R&D outcomes are stochastic, there is an additional reason for the government to differentiate. Since the firm's net profit in (6) is convex in T, offering a convex schedule of patent lives will increase the incentives to undertake R&D. In the extreme case where all firms have the same R&D productivity that they learn only after innovating (i.e. R&D outcomes are purely stochastic), it may be optimal to offer a zero patent for low values of θ and an infinite life for high values. The details of the analysis are available on request.

[16] The intuition also applies to the case of "pure obsolescence", where there is some probability each period that the firm's returns from the invention fall to zero. See Lanjouw (1998) for parametric estimates of obsolescence, using patent renewal data.

[17] A more general set-up would model second period returns π_2 as correlated with π according to a conditional distribution function $G(\pi_2, \pi)$ which satisfies the first-order stochastic dominance, $G_\pi < 0$. Then the optimal patent schedule and fees would depend on the nature of the learning process, as reflected in the function G (Cornelli and Schankerman, 1996). A general treatment, applied to the case of pollution permits, can be found in Laffont and Tirole, 1994.

The Revelation Principle also holds in this setting (Townsend, 1982), so we can focus on the direct mechanism. The government sets up a mechanism in which each firm announces $\hat{\theta}$ in period 1 and $\hat{\pi}_2$ in period 2. The mechanism specifies a fee $f_1(\hat{\theta})$ for a patent length $T_1(\hat{\theta})$ in the first period, and a fee $f_2(\hat{\theta}, \hat{\pi}_2)$ for length $T_2(\hat{\theta}, \hat{\pi}_2)$ in the second period. For simplicity we restrict the mechanism so that $T_1(\hat{\theta}) \leq \tau$: the firm can either choose a length shorter than τ—in which case renewal never arises—or it has to choose whether to renew at date τ.

Rather than fully characterise the optimal mechanism for this special case, we focus on whether it is better for the government to allow firms to abandon the patent rather than to require the full patent fees to be paid up-front.

PROPOSITION 4: *It is optimal for the government to use a patent renewal scheme rather than an ex-ante payment scheme*

Proof: See Appendix A2.

The intuition for this result is that expected profits are *convex* in the level of *ex post* profit (i.e. in the random component associated with learning), so the risk neutral firm prefers the option of taking out a longer patent life if *ex post* profits turn out to be high. Profits are convex because the optimal patent scheme involves giving longer patent lives to more profitable inventions (for incentive compatibility).

6.3. Appropriability Environment

The optimal patent policy depends on the appropriability environment, including patent licensing rules and other aspects of antitrust policy (Gallini and Trebilcock, 1995; Gilbert and Shapiro, 1997). The reason is that the flow welfare and deadweight loss will depend on the degree of appropriability, as well as the size of the invention. We briefly sketch how the model can be extended to incorporate this interaction.

Let λ denote the degree of appropriability (the fraction of the invention's social benefits received by the firm: $\frac{\pi}{B}$). As in Section 2, we can invert $\pi = \pi(\lambda, \mu)$ and write flow welfare and deadweight loss as $W(\pi, \lambda)$ and $D(\pi, \lambda)$. To incorporate appropriability in the R&D process, we write $\pi = h(z, \lambda, \theta)$, where all derivatives are positive. Note that total social benefits $B(\mu)$ do not depend on λ, so that $W_\lambda = -D_\lambda$.

It is important to distinguish between two types (or sources) of appropriability, which differ in terms of how they are likely to affect flow welfare and deadweight loss. The first arises from factors that enhance the patentee's ability to appropriate consumer surplus (e.g. more freedom to price discriminate). This gives the firm greater ability to extract *inframarginal* rents and should be associated with *increased* flow welfare and *decreased* deadweight loss (given the size of the invention). The second type arises from greater market power due, for example, to it being more difficult to imitate the invention either for legal or technical reasons. This gives the firm greater rent by *shifting the margin* and should be associated with *decreased* flow welfare and *increased* deadweight loss.

Once the welfare and deadweight loss functions are specified, the optimal patent policy, obtained using the methods in this paper, will be a function of the appropriability parameter λ. It is straightforward to derive the comparative statics of a change in λ on the optimal patent menu. The implications will depend, among other things, on whether the change in appropriability is associated with more lenient antitrust or greater market power (i.e., on how it affects flow welfare and deadweight loss). Thus, while no general conclusions can be reached without further specification, the framework can be used to study this issue.

7. Conclusions

This paper shows how patent renewal fees can be used to differentiate patent lives and thereby solve problems of asymmetric information between R&D firms and policymakers. This allows the government to provide R&D incentives more efficiently. Our approach emphasizes how heterogeneity among firms is crucial in determining the optimal use of patent fees. We also illustrate with simulation analysis how to implement the optimal patent mechanism and compare the key features of the simulated optimal mechanism with existing patent schemes.

The model in this paper has been deliberately simplified, but the reasoning underlying the derivation of the optimally differentiated patent mechanism should apply to richer specifications. The application of optimal regulation under various forms of asymmetric information can be extended to policy design in many other contexts (see Laffont and Tirole, 1994, for an application to pollution regulation). The research can be extended in several useful directions. First, post-patent learning is an empirically important feature, and more detailed study of this issue is warranted. The simulation analysis in Section 6 can be extended to study how post-patent learning and R&D uncertainty affect the optimal patent mechanism. Second, appropriability conditions are an important determinant of R&D incentives and hence of the optimal patent mechanism, and require further modelling. Finally, the analysis should be extended to a dynamic framework with sequential innovative activity and strategic interaction among firms.

APPENDIX

A1. Proof of Proposition 3

(i) Define

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

By the envelope theorem

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

Reintegrating this equation we obtain

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

where K is a constant of integration. Given that the fees do not enter the government's objective function, K is not uniquely defined. However, the minimum K will satisfy the individual rationality constraint with strict equality: $U(0) = 0$. If we set $K = 0$ and equate $U(\theta)$ to $U(\theta, \hat{\theta})$, defined in equation (9) and evaluated at $\hat{\theta} = \theta$, we obtain equation (14). This yields the maximum fees consistent with individual rationality and incentive compatibility in this model.

(ii) Maximizing (10) pointwise on $[0, \bar{\theta}]$ yields the first order conditions (13).

(iii) The sufficient conditions are given by the second order conditions for the maximization of (10) and for the incentive compatibility to hold. The first of these is given by

$$\frac{1}{r}\left[W_\pi + e^{-rT}D_\pi\right]H_{TT} + \frac{1}{r}\left[W_{\pi\pi} + e^{-rT}D_{\pi\pi}\right]H_T^2 - (z_T^*)^2 - z^* z_{TT}^* + re^{-rT}D(\pi) - 2e^{-rT}D_\pi H_T \leq 0$$

The second order condition for incentive compatibility—the maximization in (12)—is

$$U_{\theta\hat{\theta}} = T_\theta(\theta)[h_\theta e^{-rT} + A(\theta)h_{\theta z} z_T^*] \geq 0.$$

This implies that $T_\theta(\theta) \geq 0$. To guarantee $T_\theta(\theta) > 0$, note that by implicit function theorem $T_\theta(\theta)$ has the same sign as $R_\theta(\theta)$, so the sufficient condition will be satisfied if $R_\theta(\theta) > 0$.

A2. Proof of Proposition 4

A firm chooses z to maximize net profits. Consider the decision problem for a firm with a θ high enough so that it survives at least to τ (the other case is identical to the analysis in Section 4):

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

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

This yields

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

The government maximizes its objective function subject to the individual rationality constraint of the first and second period

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

and

$$\frac{\pi_2}{r}\left[e^{-r\tau} - e^{-rT_2(\theta,\pi_2)}\right] - f_2(\theta,\pi_2) \geq 0. \tag{A.3}$$

There is also the first period incentive compatibility constraint, which requires that the firm does not misrepresent θ given that it anticipates reporting truthfully in period 2,

$$\theta \equiv \text{argmax}_{\hat{\theta}} EU(\theta,\hat{\theta}) \tag{A.4}$$

and the second period incentive compatibility constraints:

$$f_2(\hat{\theta},2\pi) \geq f_2(\hat{\theta},0) \ \ \forall\hat{\theta} \tag{A.5}$$

and

$$\frac{2\pi}{r}\left[e^{-rT_2(\hat{\theta},0)} - e^{-rT_2(\hat{\theta},2\pi)}\right] \geq f_2(\hat{\theta},2\pi) - f_2(\hat{\theta},0) \ \ \forall\hat{\theta} \tag{A.6}$$

For $\pi_2 = 0$, equations (A.2) and (A.3), and the fact that $T_2(\hat{\theta},0)$ does not have any incentive effect—as it is clear from equation (A.1)—imply that $T_2(\hat{\theta},0) = 0$ and $f_2(\hat{\theta},0) = 0$. That is, implementation of this mechanism requires that the government leave firms the option to abandon their patent in case the invention turns out to be low-valued. Note that this proof relies only on the individual rationality constraints for the two periods. The incentive compatibility constraints are given only for completeness. ■

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.

CORNELLI, F. AND SCHANKERMAN, M. 1996, "Optimal Patent Renewals," *LSE working paper*, EI/13.

DE LAAT, E. 1997, "Patent Policy and the Timing of Imitation", *mimeo*, Tinbergen Institute, Erasmus University, Rotterdam.

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.

GILBERT, R. AND SHAPIRO, C. 1997, "Antitrust Issues in the Licensing of Intellectual Property: The Nine No-No's Meet the Nineties," *Brookings Microeconomic Papers*, pp. 283-349.

GREEN, J. AND S. SCOTCHMER 1995, "On the Division of Profit in Sequential Innovation", *Rand Journal of Economics*, **26**, n.1.

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., R. HENDERSON AND M. TRAJTENBERG 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. 1998, "Patent Value in the Shadow of Infringement: Simulation Estimations of Patent Value," *Review of Economic Studies*, (forthcoming).

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.

O' DONOGHUE, T., SCOTCHMER, S. AND J.-F. THISSE 1998, "Patent Breadth, Patent Life and the Pace of Technological Progress", *Journal of Economics and Management Strategy*, **7**, no.1, pp. 1-32.

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

SCHANKERMAN, M. 1998, "How valuable is patent protection? Estimates by Technology Field" *Rand Journal of Economics*, **29**, no.1, pp. 77-107.

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.

SCOTCHMER, S. 1996, "Protecting Early Innovators: Should Second Generation Products be Patentable?", *Rand Journal of Economics*, **27**, no.2, pp. 322-331.

SCOTCHMER, S. 1997, "On the Optimality of the Patent System", Working Paper # 236, Graduate School of Public Policy, Berkeley.

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.