Do patents over-compensate innovators?

Vincenzo Denicolò
Università di Bologna

1. INTRODUCTION

How much patent protection should be accorded inventors and innovators? Is protection too strong or too weak? Over the years, these questions have spawned heated controversies. Machlup and Penrose (1950) vividly recount the XIX century debate that led such countries as Switzerland and the Netherlands to delay the adoption of patent legislation and even to repeal their patent laws for a time. In the 1970s and the 1980s, by contrast, stronger patent protection was widely advocated as a remedy for the productivity slowdown. Now, after more than two decades in which patent protection has been on the rise, economists are increasingly voicing concerns that today patent protection may be excessive. Proposals are again being advanced to weaken patent protection or abolish it outright (e.g., Boldrin and Levine, 2002), explicitly or implicitly arguing that patents substantially over-reward innovators with respect to the socially optimal level.

Is this claim supported by economic analysis? Textbook models suggest that the optimal level of patent protection is the outcome of a trade-off between rapid innovation and static allocative distortion (Nordhaus, 1969). This trade-off is well understood in

This article is a revision of the paper presented at the 46th Panel Meeting of *Economic Policy* in Frankfurt. I am grateful to three anonymous referees, my discussants Pierre-Yves Geoffard and Josef Zweimuller, editor Paul Seabright, Philippe Aghion, Alan Drazen, Luigi Guiso and other panel members for helpful comments. I also benefited from comments by Laura Bottazzi, Lorenzo Casaburi, Emmanuel Duguet, Luigi Franzoni, Christine Halmenschlager, Andrea Ichino, Mark Lemley, Steven Martin, Enrico Santarelli and seminar audiences at Kiel, Toulouse, Nancy, Paris, Milan, and Trento. The usual disclaimer applies. Financial support from the University of Bologna ("Progetto d'Ateneo") is gratefully acknowledged.

theory, but in practice its optimal resolution has proved intractable. More recent and sophisticated theoretical analyses have identified new, subtler effects of patent protection (see Scotchmer, 2004, for an excellent survey), but it is clear that theory alone cannot provide answers.

The empirical literature has attacked the problem by trying to measure the quantitative impact of actual changes in patent policy.[1] However, robust empirical findings have proven to be disappointingly elusive (see e.g. Lerner, 2002). Perhaps past reforms have not been radical enough or geographically widespread enough to have a sizeable impact on the incentives to innovate,[2] and many confounding factors make it difficult to distinguish the effects of marginal policy changes.

This paper is intended to provide a firmer empirical basis for analysis. To this end it links the theoretical models of optimal patent design to the empirical findings of the innovation production function literature initiated by Griliches and Pakes (1980). This literature assesses how R&D expenditures generate innovative technological knowledge. Overcoming a good many difficulties, a number of empirical studies have found a strong relationship between R&D spending and inventions, with an elasticity as high as 0.5 or more.

Building on these empirical findings, I reformulate the theory to highlight the relationship between the level of patent protection and the elasticity of the supply of inventions. In my baseline model of stand-alone innovations, this elasticity is the sole determinant of the optimal degree of protection. Thus, the estimated elasticity offers a benchmark for assessing whether the level of patent protection is too high, too low, or near optimal. I then provide a preliminary assessment, which is that the preponderance of the empirical evidence suggests that in the aggregate patents do not over-compensate innovators. The evidence also suggests that the level of patent protection matters: for reasonable values of the elasticity and the current level of protection, raising protection to the optimal level might even double the net social benefits from innovation.

Moving beyond the baseline model, additional effects come into play. I analyze a number of extensions to determine the sign, if not the magnitude, of the additional effects. The exercise is inevitably tentative, but the outcome suggests that my preliminary assessment is not severely biased against the over-reward hypothesis. For a more precise assessment, however, more empirical research is needed.

The paper is organized as follows. The next section briefly explains how the patent system works. Section 3 develops the baseline model. Section 4 reviews the empirical literature and assesses the current level of patent protection. Section 5 analyzes a number of extensions. Section 6 summarizes and concludes the paper.

---

[1] There have also been many studies of the private and social returns to R&D, which generally find that the social returns are much larger than private returns. However, a comparison between the private and social rates of return does not directly test the over- and under-reward hypotheses. Because patents are distortionary, it is socially costly to use them as incentives for investment in research. Thus, optimally the private returns from innovation should fall short of the social returns, and an appropriate benchmark for the ratio of the two is necessarily lower than one. After deriving such a benchmark, I discuss the empirical literature more fully later on.

[2] While innovation is typically a global phenomenon, patents are national rights. A firm seeking worldwide patent protection must apply in each country separately under national law (the European patent being a notable, but partial exception). As a consequence, patent policy changes at the national level are likely to have a relatively small impact on the overall incentive to innovate.

2. BACKGROUND

A patent gives an inventor the exclusive right to manufacture, use or sell the invention for a fixed period of time. That is, it grants a legal monopoly. Patents are issued by patent offices, which screen applications to ascertain whether they meet the patentability requirements. If a patent is contested, however, its validity is eventually decided by the courts.

There are four main requirements for patentability: novelty (the innovation must not already be in the public domain), non-obviousness (some minimal inventive step is required), utility (innovative knowledge must have potential commercial applications: laws of nature cannot be patented), and subject matter (determining which technological fields are eligible to patent protection).

If an innovation is patentable, the boundaries of the monopoly are set by the patent claims. If the monopoly was confined literally to the device or product disclosed in the patent specification, the patent would be practically useless. The patent claims are agreed between inventor and patent office, often after a long bargaining process, but claims may subsequently be struck down by the courts. A patent is infringed when another device or product violates at least one of its claims; and the "doctrine of equivalents" allows a court to find infringement even where the contested device or product does not fall within the literal scope of the patent's claims.

To be of any economic value, patent rights must be enforced: as has aptly been said, a patent is just a "ticket to sue." If an infringement is found, the courts may award the patent-holder money damages or issue injunctions. Money damages can be calculated by various criteria, most notably "lost profit" and "reasonable royalty." Generally, however, the most powerful remedy is an injunction, and this is what most patent-holders seek.

An innovation may be patentable and yet infringe on another patent. Most commonly this happens when innovation is cumulative, with a second-generation technology improving upon a preceding innovation. In these cases, the holder of the dominant patent (i.e., the one covering the first-generation technology) can continue to use his innovation freely, but neither he nor the holder of the subservient patent can use the second-generation technology without a licensing agreement. This legal solution can be used to protect the original innovator against competition from minor improvements and to allow him to share the profits from the second-generation technology to which he opened the way.

2.1. The profit ratio

In the light of this stylized description of the patent system, what is the level of patent protection and how would we measure it? If the invention is commercially valuable,[3] the patent-holder will generally be able to make a profit by using the innovation himself,

[3] It is a well established empirical fact that the distribution of patent values is highly skewed: most patents are practically valueless, and only a tiny fraction are highly profitable (Harhoff *et al.*, 1999).

licensing it to others, or both. But since patent protection is limited, the patent-holder's profit will be constrained below the maximum hypothetical level. One can then measure the degree of protection as a profit ratio, namely, the ratio of the discounted profit actually captured to the hypothetical discounted profit the patent-holder would get with infinitely long, complete monopoly control over the innovation.

What determines the profit ratio? Many economists who are not intellectual property experts believe that the main reason why patent protection is limited is the finiteness of the statutory patent term – currently 20 years in most countries. In fact effective patent life is frequently shorter than 20 years, because patents are often obtained long before products are actually marketed. But patented innovative knowledge can often be exploited commercially by others before the patent expires. This is so for a variety of reasons. First, the enforcement of patent rights is largely incomplete. Second, "inventing around" a patent is a strategy encouraged by the law and routinely used by competitors to reduce the patent-holder's competitive advantage. Third, patent protection is limited geographically, and in spite of the TRIPs agreement it is still practically absent in many countries.[4] Last but not least, subsequent improvements building on an original innovation may make the innovation itself at least partially obsolete. All of the foregoing factors concur to determine what patent economics jargon calls "breadth." The length and breadth of patent protection jointly determine the profit ratio.

It is therefore apparent that for any given innovation the profit ratio depends on a number of technological details and random events, but it also depends on legal rules and the decisions of governments, patent offices, and the courts. For example, the scope of the claims approved by patent offices and upheld by the courts determines how difficult it is to invent around a patent and how large an improvement must be in order not to be ruled in infringement of the original patent. Money damages awarded in case of infringement provide direct compensation for the patent-holder and may have an indirect deterrence effect. Policy rules governing the parallel trade of patented products determine the extent to which patent-holders can engage in price discrimination. Competition policy determines whether or not patent-holders can leverage their market power by tying other products to the patented product, and so on. And, of course, the profit ratio depends on patent life.

One approach would be to introduce all these and other specific policy tools into a model and to examine the optimal choice of each. Indeed, many important questions related to patent policy turn on the choice of the most appropriate combination of policy tools. Here, however, I take an alternative approach, abstracting from the complicated institutional details of patent law and asking a more basic question: how large should the pie granted to innovators be? The profit ratio summarizes many policy choices in a single index and thus is a convenient tool for addressing this question.

---

[4] Negotiated under the Uruguay round of the GATT, the TRIPs agreement sets compulsory minimum standards for patent protection for WTO member countries. McCalman (2001) analyzes the effects of national patent reforms enacted to comply with the TRIPs agreement.

2.2. The division of profit

When innovation is cumulative, or several innovative components must be assembled to operate a new technology, patent policy must determine not only the size of the pie, but also how it should be shared among all the firms that contributed to the discovery. If some innovators get more than their fair share and others less, incentives to innovate may be seriously impaired. Whose profits are then to be included in the profit ratio?

In this paper I abstract from the issue of the division of profit and focus on the *aggregate* profit ratio. Although this is a useful starting point for the analysis, the reader should be advised that certain policy reforms that are described as moves towards stronger patent protection actually affect only the division of profit and may not raise the aggregate profit ratio at all. For example, recently the American patent system has experienced an unprecedented broadening of patentability. There are now thousands of patents for living organisms, genetic sequences, software programs, business methods etc., none of which would have been patentable in the 1970s.[5] Have these patents resulted in a higher aggregate profit ratio? Not necessarily. Let us consider, for example, gene patents. Often, decoding genetic sequences is instrumental to the search for new drugs. Since new drugs have long been patentable, however, the patentability of gene sequences will not necessarily increase the drug company's profits. On the contrary, making genetic sequences patentable may create a problem of Cournot complements,[6] or it may increase transaction costs and so lower the aggregate profit ratio in the industry. In addition, and perhaps most important, making genetic sequences patentable may affect the division of profit between upstream and downstream firms, changing the incentives to conduct upstream and downstream research. The final effect on innovative activity is most uncertain.[7]

3. THE BASELINE MODEL

The idea behind many models of optimal patent protection is very simple: an innovation is in prospect but is costly to achieve. Lacking patent protection, nobody would invest in R&D because innovative knowledge is non-excludable and R&D expenditures could not be recouped. Patents alleviate this appropriation problem, but they generate market power and thus impose deadweight losses. This creates a trade-off between innovation and monopoly distortions.

There have been many studies of optimal patent policy along these lines: see e.g. Duffy (2005) and the literature cited therein. Here I reformulate the analysis to produce a parsimonious characterization of the optimal level of patent protection that enables us to

[5] Things are somewhat different in Europe, where business methods and software patents are still controversial and gene patents are much harder to uphold than in the US.

[6] When several separate monopolists supply complementary components of the final product and price each component non-cooperatively, the price to the final consumer may exceed the monopoly price, reducing aggregate profits below monopoly profits (Shapiro, 2001).

[7] Similar considerations may apply to software programs and may help explain why the introduction of software patents in the US does not seem to have boosted innovation (Bessen and Hunt, 2007).

exploit the empirical findings of the literature on the innovation production function. While none of the assumptions of the baseline model are new, when brought together they yield a strikingly simple result: the profit ratio should just equal the elasticity of the supply of inventions.

3.1. Model outline

The model posits that a large number of innovations are in prospect and that their nature and size are fixed. For a representative innovation, π denotes the (flow) profit that would be earned by a fully protected innovator, D the corresponding deadweight loss, and V the social value of the innovation once it is in the public domain. We can write $V = \pi + D + CS$, where CS is a residual. In the case of product innovation described in Box 1, for example, CS is consumer surplus, but it may also capture technological spillovers and other positive externalities. To keep the analysis as general as possible, π, D, and CS are treated as fixed parameters, without any specifics on the nature of the innovation, demand, and the competition in the product market. For simplicity the environment is assumed stationary. Hence, the social value of the innovation is $v^S = V/r$, where r is the social – and private – discount rate.

Patent policy determines what fraction of π/r patent-holders can actually capture. I focus on two policy variables, namely the length and the breadth of patent protection. The meaning of length is straightforward. Given a patent life of T years, it is convenient to define the "normalized" length as $z = 1 - e^{-rT}$. The variable z can then be interpreted as a profit ratio: it is the present discounted value of a constant annuity over the patent lifetime T as a fraction of the present discounted value of a perpetual constant annuity. To capture the fact that patents are limited in breadth, following Gilbert and Shapiro (1990) I assume that the innovator earns a fraction β of the flow profit π while the patent is in force; the remainder $(1-\beta)$ spills over to somebody else (consumers or other firms). Thus, when patent protection is limited in scope and over time the innovator's discounted profit is $v^P = \beta z \pi / r$: normalized length z and breadth β combine into a single variable, the profit ratio βz, which indexes the level of patent protection. For any given prospective innovation, the profit ratio determines the incentive to innovate.

The cornerstone of the model is the innovation production function that links the incentive to innovate to the supply of inventions. I assume that the innovation can be achieved instantaneously with an aggregate probability x at a cost $\alpha c(x)$ with $c(0) = 0$. The aggregate R&D expenditure function $c(x)$ is increasing and convex: $c'(x) > 0$ and $c''(x) > 0$, and α is a positive parameter that measures the difficulty of achieving the innovation. Box 2 and Appendix A show that this simple formulation can be regarded as a reduced form of many seemingly different models of investment in research that are used in the literature, including the Poisson model. The convexity assumption captures the notion that the production of innovative knowledge requires at least one input whose supply is fixed, such as talent or the set of good ideas at any given point in time. For simplicity, parameter α is assumed to be large enough that all equilibrium conditions considered below deliver interior solutions.

Box 1. Drastic and non-drastic innovations

Figure 1 depicts a radical product innovation. Before the innovation the good cannot be produced (i.e., its unit production cost, c_0, is higher than the intercept of demand). With the innovation, the patent-holder (or the licensee) charges the monopoly price p_M for as long as the patent runs. In this phase the flow social benefit from the innovation is $\pi + CS$. When the patent lapses, the price falls to marginal cost c_1 and society nets a flow benefit of $V = \pi + D + CS$.

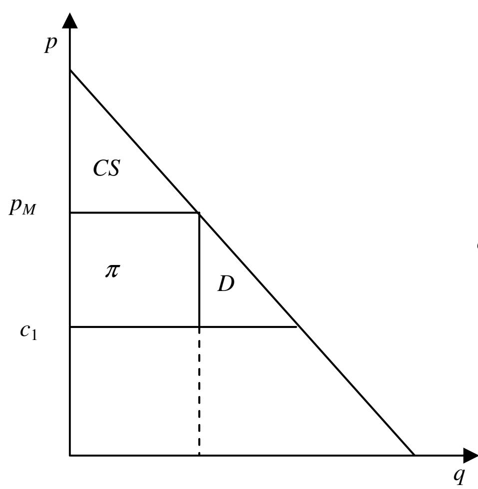

Figure 1. Product innovation

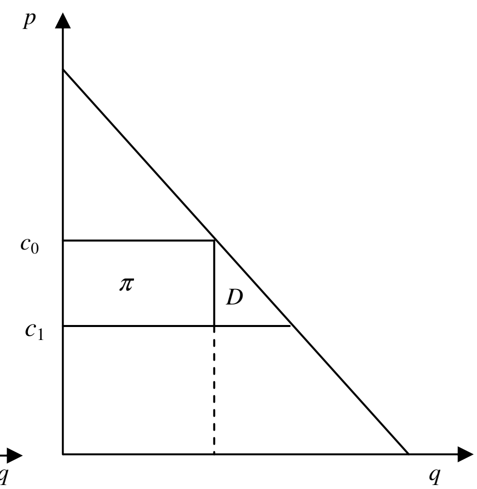

Figure 2. Non-drastic innovation

Figure 2 depicts the case of a non-drastic innovation. Here the innovation reduces the unit production cost from c_0 to $c_1 < c_0$; now p_M (not shown in the figure) is greater than c_0. Assuming that the market is perfectly competitive before the innovation and that the innovator engages in limit pricing, the price will remain at the pre-innovation level c_0 until the innovation falls into the public domain. When the patent expires, the price falls to marginal cost c_1 and society obtains the full flow social value of the innovation. Here $CS = 0$ (CS denotes the *increase* in consumer surplus due to the innovation) and therefore $V = \pi + D$. This is the case focused on in the baseline model; buyers do not benefit from the innovation until the patent expires.

Inverting the R&D expenditure function, we get an innovation production function that expresses the probability of success, x, as a function of R&D expeditures. The elasticity of this innovation production function is $\eta = c(x)/xc'(x)$. With a large number of potential inventions, the supply is proportional to the probability of success, so η is precisely the elasticity of the supply of inventions that has been estimated in the empirical literature. It is necessarily less than one given that the function $c(x)$ is convex and $c(0) = 0$.

Finally, in the baseline model $CS = 0$ and there is free entry into research, which will drive the expected net profit of innovators to zero. With these assumptions, the baseline model consists of two simple equations. The first equation is the zero-profit condition in the research industry:

$$xv^P - \alpha c(x) = 0 \tag{1}$$

This equation implicitly gives the equilibrium level of investment in research as a function of the profit ratio. It says that with free entry, all of the expected profit from the innovation is invested in research.

Box 2. An example

Assume that each research firm can engage in an indivisible research project and that different firms' projects are uncorrelated. Each project succeeds with probability p, with $0 < p < 1$, and costs q. Then, if n projects are run, the aggregate probability of success is the probability that at least one project will succeed, i.e., $x = 1 - (1- p)^n$, and the aggregate R&D expenditure is nq. Since $n = \log(1- x)/\log(1- p)$, the aggregate R&D expenditure can be rewritten as $nq = \alpha c(x)$ where $\alpha = -q/\log(1-p)$ and $c(x) = -\log(1-x)$. It is easy to check that this function is increasing and convex, and satisfies the condition $c(0) = 0$. Next, I suppose that if two or more projects succeed simultaneously, one project will be selected at random and the patent granted to the firm running that project. Then the expected profit from running a project is $(x/n)v^P - q$. The zero-profit condition requires that $(x/n)v^P - q = 0$, or equivalently $xv^P - nq = 0$, whence it is clear that the zero-profit condition reduces to (1). Appendix A provides more examples like this, including the standard model in which the timing of the innovation is uncertain and follows a Poisson process.

The second equation of the baseline model is the social welfare function to be maximized. I use the standard definition of social welfare employed in most partial equilibrium analyses, i.e., the discounted sum of profit and consumer surplus: [8]

$$W = x\left[z\frac{\pi + (1-\beta)D}{r} + (1-z)\frac{(\pi + D)}{r} \right] - \alpha c(x) \tag{2}$$
$$= x(1-\beta z)v^S$$

[8] Some papers analyze optimal patent protection in a general equilibrium framework: see e.g. O'Donoghue and Zweimuller (2001), Grossman and Lai (2004) and Acemoglu and Akcigit (2006). Qualitatively, these models exhibit the same trade-off of innovation and monopoly distortions as their partial equilibrium counterparts. Relatively little is known on the magnitude of the general equilibrium effects. One notable exception is Saint-Paul (2004), who asks whether distributional concerns might not reduce the optimal strength of intellectual property protection and finds that any such effect is small. More research along these lines is needed, however.

Let us explicate the meaning of this equation. The innovation is achieved with probability x. Conditional on success, the flow social value of the innovation is $[\pi + (1 - \beta)D]$ until the patent expires, i.e., for a normalized time length z, and $(\pi + D)$ thereafter, i.e., for the remaining time $1 - z$. (Here I assume that when breadth is limited, the deadweight loss is proportionally reduced, i.e., it is βD. The consequences of relaxing this and other assumptions are discussed later on.) Finally, to obtain net social welfare, R&D costs are subtracted.

The second line of equation (2) can be obtained by plugging the zero-profit condition into the social welfare function. It says that net social welfare is a fraction of the expected social value of the innovation, xv^S, and neatly illustrates the trade-off between innovation and monopoly: a rise in the profit ratio βz increases the probability of success x but decreases the fraction $(1-\beta z)$ of the social value society nets upon discovery.

3.2. The elasticity rule

The policy-maker chooses the length and the breadth of protection, z and β, so as to maximize social welfare (2) considering that x is implicitly given by the zero-profit condition (1). The solution yields the following simple "elasticity rule:"

$$\beta z = \eta \tag{3}$$

which says that the profit ratio should be equal to the elasticity of the supply of inventions (see Appendix B for a proof).[9]

Some insight can be gotten by noting that at the optimum a rise in the profit ratio must increase the supply of inventions by the same percentage amount as it lowers the degree to which society benefits from the innovation, $(1-\beta z)$. Now, a 1 percent rise in the profit ratio βz increases the innovator's expected profit by 1 percent. By the zero-profit condition, this increases R&D expenditure by 1 percent, which in turn increases the supply of inventions by η percent. This, however, is only the direct effect. If the probability of success increases by η percent, the expected profit increases by the same amount, and by the zero-profit condition so does R&D expenditure. This further raises the supply of inventions, with a multiplicative effect that is similar to a Keynesian multiplier. The total effect of a 1 percent increase in the profit ratio is therefore a percentage increase in the supply of inventions equal to $\eta + \eta^2 + \eta^3 + ... = \eta/(1 - \eta)$. Since the elasticity of $(1-\beta z)$ with respect to the profit ratio is simply $\beta z/(1-\beta z)$, the policymaker must equate $\eta/(1 - \eta)$ and $\beta z/(1-\beta z)$. This immediately yields the elasticity rule.

[9] A number of works have analyzed optimal patent life, but so far the role of the elasticity of the supply of inventions has tended to be overlooked. Some papers implicitly assume specific numerical values of the elasticity (for instance, with a quadratic specification of the R&D cost function $c(x)$ the elasticity is ½); others use specific functional forms that make the elasticity itself depend on the level of R&D spending and other parameters, which obscures the economic interpretation of the results.

The elasticity rule (3) implies that the optimal level of patent protection does not depend directly on various innovation-specific parameters, such as π, D and α. This can be explained intuitively. First, an increase in the size of the innovation or in the size of the market, as measured by $\pi + D$, increases the size of the investment in research that society ought to make to achieve that innovation. For any given level of the profit ratio, however, a rise in π raises the private value of the innovation and so automatically raises the equilibrium level of R&D investment. As a result, there is no need to adjust the strength of protection. For similar reasons, the optimal level of patent protection does not depend directly on the difficulty of achieving the innovation as captured by the shift parameter α. Finally, let us consider the deadweight loss to profit ratio D/π. This ratio measures how costly it is to provide incentives for R&D investment through patents, and so determines the terms of the trade-off between innovation and monopoly distortions. It may seem surprising that it does not affect the optimal level of patent protection directly. The intuition is that until the patent expires, the share β of both D and π is lost (D is lost due to the static inefficiency of supra-competitive pricing and π is dissipated in the patent race due to the free entry assumption) and so society enjoys only a share $(1 - \beta)$ of $\pi + D$; when the patent expires, there is no longer any deadweight loss or any profit and society enjoys the entire social value of the innovation, V.[10]

However, parameters such as π, D and α do determine the equilibrium level of R&D expenditure, i.e., the point on the innovation production function. For example, other things being equal, a decline in D/π increases the equilibrium investment in R&D, hence the number of inventions and social welfare. Moreover, if the elasticity of the supply of inventions changes with the level of investment in research, any variable that affects R&D expenditure impacts the optimal level of patent protection indirectly, via the elasticity η.

The fact that the optimal strength of protection does not depend directly on many innovation-specific parameters allows for an immediate extension of the elasticity rule to the case of heterogeneous innovations. With a distribution of innovations that differ in value or R&D cost, equation (3) continues to hold if the elasticity η is the same for all innovations. More generally, the optimal strength of protection is a weighted average of the individual elasticities, the weights reflecting the values of different innovations. Moreover, heterogeneity increases the optimal level of protection, in that a mean-preserving spread in the distribution of the individual elasticities raises the optimal profit ratio (see Appendix B for details).

[10] It may seem particularly surprising that the optimal profit ratio is less than 1 even when $D = 0$. When $D = 0$, in the baseline model there are no social costs associated with patent protection, so it would seem that setting $\beta z = 1$ is desirable. However, recall that with competition in research the winner-takes-all effect leads to over-investment. Reducing the profit ratio mitigates this over-investment problem. On the other hand, I show below that with monopoly in research and $D = 0$ it is indeed optimal to set $\beta z = 1$.

4. PRELIMINARY ASSESSMENT

Let us now turn from theory to empirical evidence. In this section, I first review empirical estimates of the elasticity of the supply of inventions with respect to R&D expenditure.[11] Next, I tentatively assess the value of the profit ratio in current practice. In view of the elasticity rule, the comparison between the profit ratio and the estimated elasticity of the supply of inventions provides a preliminary assessment of whether patent protection is too strong or too weak.

4.1. Empirical estimates of the elasticity

What is the most appropriate evidence for estimating the elasticity of the supply of inventions? In an ideal experiment, one would observe two groups of firms that are identical in all respects, except that one invests in R&D more for some exogenous reason. One could then calculate the elasticity of the supply of inventions by counting how many more inventions firms in the high-investment group achieve.

In practice, of course, this ideal experiment can never be run. We have data on different firms and regions in the world, but they differ in countless respects; R&D expenditures do not vary exogenously; and inventions must often be proxied by patents or patent applications. Using this data, at the very best econometricians can only approximate the ideal experiment.

A most creative approximation is Acemoglu and Linn (2004). They examine the pharmaceutical industry, where the number of newly approved drugs (not the number of new chemical substances patented, many of which fail to pass pre-clinical and clinical tests) is a fairly precise measure of the number of innovations. They do not look at R&D expenditure but measure directly the impact on innovation of changes in the size of the market (which in their sample varies exogenously for demographic reasons). They find that a 1% increase in the size of the market for pharmaceutical products raises the number of new drugs by 4% to 6%. These results imply an elasticity of the supply of inventions in the pharmaceutical sector ranging from 0.8 to 0.86.[12]

Turning to more standard evidence, a number of empirical studies have sought to estimate an innovation production function

$$P_t = f(R_t, R_{t-1}, ...) \tag{4}$$

that relates the number of patent applications at time t, P_t, to current and past R&D expenditure, R_t, R_{t-1}, … Apart from the time lags and the use of P as a proxy for x, the function $f(.)$ is just the inverse of the R&D expenditure function $\alpha c(x)$. Therefore, with a

---

[11] Although the elasticity rule also applies to copyrights, to the best of my knowledge there are no empirical estimates of the elasticity of the supply of copyrightable works.

[12] Provided that firms invest a constant share of the expected profits from the innovation in R&D (e.g., 100% with free entry), the elasticity empirically estimated by Acemoglu and Linn is equal to $\eta/(1-\eta)$. More generally, however, at the optimum their elasticity must be equal to $\beta z/(1-\beta z)$ (see Appendix B), and their results therefore directly imply that the optimal range for the profit ratio is from 0.8 to 0.86.

Cobb-Douglas specification of the innovation production function, the estimated parameters can be directly interpreted as the appropriate elasticities.

Using an NBER data set of patent applications and R&D expenditures for a sample of US manufacturing firms, Pakes and Griliches (1980, 1984) find a total elasticity (i.e., the sum of the coefficients of current and lagged R&D expenditures) of 0.61. Their analysis is extended and refined by Hausman *et al.* (1984) and Hall *et al.* (1986). These authors note that a log-linear regression model is inappropriate for firm-level data since many observations have zero patents, suggesting the use of a Poisson or a negative binomial model instead. In a sample of 128 large firms over about 10 years included in the NBER data set, Hausman *et al.* (1984) obtain an R&D elasticity of 0.87 with the Poisson and of 0.75 with the negative binomial distribution. However, their data suggests strong firm-specific effects. Accounting for these lowers the estimated elasticities: for example, with the negative binomial model the total elasticity of 0.75 falls to 0.52 with random effects and to 0.37 with fixed effects. However, this study also notes that firm-specific effects may actually reflect pre-sample R&D and concludes that while the elasticity of the supply of inventions to current R&D may be less than 0.4, the total elasticity (including lagged and pre-sample R&D) is "at least 0.55" but "could be significantly higher" (p. 933). Extending the analysis to a larger sample of 642 firms, Hall *et al.* (1986) arrive at similar conclusions. They find that there is a strong relationship between current R&D expenditure and patenting, with an elasticity of about 0.3; the contribution of observed past R&D is about 0.05; and the contribution of unobserved or pre-sample R&D is about 0.25 (p. 281). These effects add up to a total elasticity of about 0.6.

A similar pattern emerges from subsequent research that has used more general distributional assumptions (like the General Event Count model), different ways of taking account of individual heterogeneity (like GMM quasi-differenced estimation), and different data sets: see Montalvo (1997), Cincera (1997), Crépon and Duguet (1997a,b), Wang *et al.* (1998), Branstetter (2001) and Guo and Trivedi (2002). In these studies, levels estimations of the elasticity are typically large, around 0.8-1, but they diminish sharply when individual heterogeneity is taken into account in some fashion. For example, using quasi-differenced GMM estimation, Cincera (1997) finds an elasticity of 0.48 and Crépon and Duguet (1997a) as low as 0.26.

Blundell *et al.* (2002) suggest a possible resolution to this apparent inconsistency. They note that the process under study is inherently dynamic: as a matter of fact, it is often very difficult to reject the hypothesis that R&D expenditures follow a random walk. Using Monte Carlo analysis, they show that when the regressor is auto-correlated, levels estimators that do not take unobserved heterogeneity into account are upward biased while the estimators commonly used to account for firm-specific effects are severely downward biased. They propose an alternative "pre-sample mean" estimator replacing the fixed effect with the pre-sample mean of the dependent variable – here, the number of patents. This estimator performs well in the Monte Carlo simulation, and when applied to the NBER data set delivers a point estimate of the elasticity η of 0.51.

One problem with the use of patents as a proxy is that not all innovations are patented. Some recent work exploits survey data as well as patent data to model the decision to patent explicitly. Duguet and Kabla (1998) use data from the French Technological Appropriation Survey to construct a variable that measures the propensity to patent. They estimate both a reduced form model of the innovation production function and a structural model in which the dependent variables are the propensity to patent and the number of patent applications. In the reduced-form model estimation, the elasticity of the supply of inventions ranges between 0.6 and 0.7, but in the structural form it falls to 0.4-0.5. Arora *et al.* (2005) use data from the Carnegie Mellon Survey and develop a structural model in which the number of patent applications, the propensity to patent, and R&D expenditures are all endogenous. The elasticity estimate is also influenced by a qualitative index of patent effectiveness that may capture differences in the value of patents across industries. They find a point estimation of the elasticity of 0.61, both in the single equation and in the system estimates.

Another group of studies (Bottazzi and Peri, 2003, and Peri, 2005) estimate the elasticity using regional-level rather than firm-level data. The very large number of patents in each region (e.g., almost 10,000 on average in Peri, 2005) relieves the problem of variability in the value of innovations, since differences tend to be averaged out in the aggregate. Bottazzi and Peri's estimates of the elasticity vary from 0.67 to 0.96, those of Peri (2005) from 0.6 to 0.81.

Jones and Williams (1998, 2000) find a lower bound for the elasticity by an entirely different line of argument. In a model of endogenous growth, they show that the elasticity equals the social rate of return to R&D, multiplied by the share of R&D in national income and divided by the rate of total factor productivity (TFP) growth. Using a lower bound of 30% for the social returns to R&D taken from the empirical literature, a lower bound of 2.2% for the R&D share, and an estimated TFP growth rate of 1.25%, they find a lower bound for the elasticity of 0.5.

Although no single piece of evidence is compelling, as a whole it suggests that the elasticity of the supply of inventions may be prudently taken to range between 0.5 and 0.7. However, that available estimates vary so greatly also suggests that the true elasticity may vary across sectors and over time. A great deal of caution is indispensable, therefore, in drawing policy conclusions.

In particular, the elasticity may change with the level of R&D investment. Although most empirical studies use a Cobb-Douglas specification of the innovation production function, the few studies that allow for more flexible functional forms suggest that the elasticity may not be constant (Guo and Trivedi, 2002). This means that unless the economy happens to be at the social optimum the optimal profit ratio does not necessarily coincide with the observed elasticity. But the latter is a proper benchmark to test the over-compensation hypothesis: patent-holders are unlikely to be over-compensated if they get less than half – possibly even up to two thirds – the discounted profits they would get with infinitely-long, complete monopoly. I now turn to the other term of the comparison: the profit ratio.

4.2. The level of the profit ratio

The measurement of the profit ratio is a big problem: to my knowledge, there is no empirical study of βz. To address the issue, in this subsection I deal separately with the normalized length z and breadth β. In the next subsection I argue that some indirect indication of the level of the profit ratio is provided by the empirical literature that seeks to estimate the private and social rates of return to R&D.

4.2.1. Normalized length

On the face of it, length is easy to measure: in most countries, statutory patent life is 20 years from the date of filing. As we have seen, however, the effective patent life is often less than that; in pharmaceuticals, for example, it has been estimated at less than 12 years in spite of various provisions extending statutory patent life to compensate for the time lost while developing the product and obtaining regulatory approval (Grabowski and Vernon, 2000).

Another problem is the choice of the discount rate for translating effective patent life (measured in calendar time) into a profit ratio like z. In keeping with the assumption that the social and private discount rates coincide, here I abstract from risk. (In an extension developed below, I introduce a risk premium allowing the social and private discount rates to differ.) Starting from the long-run, risk-free real interest rate, to obtain a proper discount rate one must:

- Subtract the economy's rate of growth to account for the growth in profits, deadweight losses etc.
- Add the instantaneous probability of exogenous technological breakthroughs to allow for the possibility that the innovation may be superseded by subsequent technical progress.

Only social obsolescence matters here, not private depreciation. Consequently, only exogenous technical progress should be taken into account, not improvements that could not have been sought had the original innovation not been made. Although in practice it is difficult to distinguish between the two types of technical progress, conceptually they are very different. Given an original innovation (say, the antiretroviral therapy against HIV), independent technological breakthroughs (an effective vaccine) may make the original innovation socially valueless. It is therefore appropriate to augment the social discount rate by the probability of such breakthroughs. But no correction should be made to account for subsequent improvements to the original innovation (like "me-too" drugs or second-generation antiretroviral drugs), since they actually enhance the social value of the innovation on which they build. The key difference is whether the superior technology that makes the original innovation privately obsolete could or could not have been achieved in the absence of the original innovation. [13]

The difference between the long-run, risk-free real interest rate and the economy's rate of growth is small, probably not more than 1 or 2%. The magnitude of the correction to account for exogenous technical progress is more speculative. If major exogenous technological breakthroughs follow a Poisson process, an instantaneous probability of 2% means that a major breakthrough will occur every 50 years, one of 5% that a breakthrough comes every 20 years. It seems unlikely that major exogenous breakthroughs could be much more frequent.

Table 1: Normalized patent length

Interest rate	3%	5%	7%	10%
Effective patent life				
20	0.45	0.63	0.75	0.86
16	0.38	0.55	0.67	0.80
12	0.30	0.45	0.57	0.70

Source: Author's calculations

Table 1 illustrates how normalized length depends on effective patent life and the discount rate. For large but reasonable values of the discount rate, on the order of 5% to 7%, normalized length does not exceed 0.75 and may be as low as 0.5. Thus, the mere fact that patent protection is limited in time pushes the profit ratio considerably below one. However, length is arguably less important than breadth in limiting the strength of protection.

4.2.2. Breadth

Patent breadth is limited for a variety of reasons:

- Detecting infringements is difficult, especially when the potential users of the new technology are numerous. Enforcement of patent rights is costly and possibly even risky, since contested patents are frequently invalidated by the courts. Thus, some infringement is inevitably tolerated.
- To some extent, patented technological knowledge can be lawfully imitated even before the patent expires. Even with the best of the intentions, patent claims cannot be drafted so broadly as to block all possibility of imitation. Because of the inherent ambiguity of human language, too broad claims would engender the risk of inventor's getting a monopoly over something that is already known, and thus should stay in the public domain, or over something not yet invented, which would

[13] From a different perspective, the erosion of the original innovator's profit is inevitable when there is an exogenous technological breakthrough, but when innovation is cumulative it is a matter of patent policy. In the latter case, if the original innovator were granted full monopoly control over his innovation, including the power to block subsequent improvements, there would be no profit erosion: any erosion that does occur in practice is attributable to the limited scope of patent protection as captured by the variable β (Scotchmer, 1991).

stifle subsequent innovation. In practice, these risks cannot be avoided save by making some imitation lawful.

- Many developing countries have practically no patent protection. This not only deprives patent-holders of profits there, but may also reduce their profits elsewhere facilitating counterfeiting and re-imports.
- Innovators face competition from subsequent improvements. If patents guaranteed complete forward protection (i.e., all subsequent improvements were deemed infringements), the original patent-holder would be sheltered from all competition and might even be able to get a share of the profits from the improvements. In practice, however, forward protection is largely incomplete, and competition from improved versions is a major source of profit erosion.

Although it is therefore evident that in practice β is substantially lower than 1, it is hard to say by how much. Some indirect indication can be drawn by survey data. Surveys in various countries show rather conclusively that patents are usually rated as less important than secrecy and lead time for appropriating the returns to innovation (Levin *et al.*, 1987; Cohen *et al*., 2002). This evidence would be difficult to reconcile with high values of β: for instance, lead time could hardly be considered more effective than patents if the latter guaranteed an iron-clad monopoly, even a temporary one.

More direct evidence on the profit erosion process comes from two sources. In a classic study, Mansfield *et al.* (1981) found that more than 60% of all patented innovations in their sample were imitated within 4 years. This evidence does not suffice to provide a direct estimate of patent breadth. However, with reasonable ancillary assumptions it suggests that the process of inventing around a patent alone pushes β below 0.6.[14]

More recently, Lichtenberg and Philipson (2002) have focused on pharmaceuticals, distinguishing between "within-patent entry" (competition from manufacturers of generic drugs when the patent expires) and "between-patent entry" (competition from producers of competing new drugs). In other words, within-patent entry reflects the limitation in patent protection in length, whereas between-patent entry captures both the limitation in breadth and exogenous technical progress. The authors find that between-patent entry has at least twice as large an effect on discounted sales as within-patent entry. Thus, if just half the between-patent competition is attributable to the fact that

---

[14] Suppose that the stochastic event of successful imitation follows a Poisson process with parameter μ. Here μ is the probability of an innovation being imitated in the current year, conditional on no imitation to date. Mansfield *et al.*'s data implies that μ = 0.23. (More recent work by Agarwal and Gort, 2001, suggests that the speed of imitation, if anything, has increased over time.) Next, suppose that "imitation" halves the innovator's profit (the patent-holder must share the market with the imitator). Then, the profit ratio is

$$r\int_0^T e^{-rt} e^{-\mu t}\left[1+\mu\int_t^T \frac{1}{2} e^{-rs}\, ds\right] dt$$

With a patent life of 20 years and a discount rate of 5% (but the calculation is not very sensitive to the choice of these parameter values), the profit ratio is 0.37, or about 58% of the profit ratio that corresponds to T = 20 and β = 1, which is 0.63 (see Table 1). This is probably an optimistic guess, however. Generally speaking, competition from imitators lowers industry profits and there can also be multiple imitations, which implies that imitation may more than halve the patent-holder's profit.

patent protection is limited in breadth,[15] Lichtenberg and Philipson's finding would imply that breadth is more important than length in reducing the patent-holder's ability to appropriate the returns. With an effective patent life of 12 years and a discount rate of 5% to 7%%, in the pharmaceutical sector normalized length z is around 0.5, suggesting that β is unlikely to exceed 0.5 either.[16] Note that patent protection is notoriously considered to be more important in the pharmaceutical sector than in many others, in spite of relatively short effective patent length (Cohen *et al.*, 2000). This suggests, conversely, that in this industry the breadth of protection is particularly large.

4.2.3. Summing up

The empirical estimates to date indicate that the elasticity of the supply of inventions ranges from 0.5 to 0.7. In practice, normalized patent length is probably below 0.75 and the breadth of patent protection is unlikely to exceed 0.5. Therefore, it seems that the profit ratio for the representative patent-holder is less than half, and perhaps even a third may well be overoptimistic. Although these estimations should be interpreted with caution, due to the paucity and noisiness of the data, they indicate that the evidence available does not support the over-compensation hypothesis.

4.3. The private and social rates of return to R&D

A complementary perspective is offered by the substantial empirical literature that has sought to estimate the private and social returns to R&D. In general, this literature finds that the social returns are much larger than the private returns. Almost all studies find that the private rate of returns is less than half the social rate, and most studies find that it is less than one third (see Hall, 1996, Jones and Williams, 1998 and Griffith, 2000 for surveys). According to Baumol (2002), one fifth is a reasonable guess.

Of course, the ratio between the private and social returns to R&D differs from the profit ratio: technological spillovers, consumer surplus and other positive externalities associated with innovative activity would make the social returns exceed private returns even if the profit ratio were equal to 1. To exploit the findings of this empirical literature, one must therefore relax the assumption $CS = 0$. Appendix B provides such an extension, deriving a lower bound for the ratio between the private and social rates of return to R&D. The following inequality is a necessary, but not sufficient, condition for innovators to be over-compensated:

$$\frac{r}{r_S} \geq \frac{\pi}{\pi + D}\eta \tag{5}$$

[15] Lichtenberg and Philipson do not unfortunately provide separate estimates of the two components of between-patent entry, namely, cumulative and exogenous technical progress, but the fact that most newly approved medicines are "me-too" drugs suggests that cumulative innovation may be more important than exogenous technical progress in explaining the profit erosion due to between-patent entry.

[16] Moreover, Lichtenberg and Philipson focus on the US and so ignore counterfeiting, a phenomenon that has been estimated to account for nearly 10% of total drug sales at the world level.

Here r is the private rate of return (with free-entry, research firms earn no extra-profits and so the private rate of return coincides with r) and r_S is the social rate of return, which is defined formally in Appendix B.

If the deadweight loss to profit ratio D/π is small, the factor that multiplies η on the right hand side of (5) is close to 1. In this case, patent-holders might be over-compensated only if the ratio between the private and social returns to R&D exceeded the elasticity of the supply of inventions. But a direct comparison of the two would then rather clearly reject the over-compensation hypothesis: the elasticity η probably exceeds 0.5, which is instead an upper bound for the ratio r/r_S.

However, even this conclusion should be interpreted with caution. Scotchmer (2004) discusses various reasons why it is not possible to relate exactly the rates of return calculated in partial equilibrium models of patent protection to estimates found in the empirical productivity literature. Moreover, the magnitude of deadweight loss to profit ratio D/π might be non-negligible. Yet, the complementary perspective offered by the productivity literature reinforces the conclusion that the over-compensation hypothesis does not seem to be supported by the data.

4.4. Potential gains from patent reforms

If the elasticity of the supply of inventions varies with the level of R&D investment, it is very hard to gauge the potential social gains from bringing patent protection to the optimal level, or even just the magnitude of the required adjustment. However, assuming a constant elasticity the baseline model (1)-(2) implies that net social welfare can be expressed as

$$W = \frac{(\pi + D)}{r}\left(\frac{\pi}{r\alpha}\right)(\beta z)^{\frac{\eta}{1-\eta}}(1-\beta z) \tag{6}$$

If η = ⅔ and βz = ⅓, equation (6) implies that the net social benefits from innovation would be doubled by raising the profit ratio from ⅓ to the optimal level of ⅔. Since the calculation is quite sensitive to the values of η and βz, however, the elasticity and the profit ratio must be measured much more precisely than has been done so far in order to assess the potential gains at all accurately. However, these rough calculations do suggest that the level of patent protection matters and the gains from bringing patent protection to the optimal level may be large.

5. EXTENSIONS

The baseline model is simplistic in many respects and many robustness checks are needed before jumping to conclusions. The restrictiveness of some assumptions is rather obvious: in real life firms need not invest in R&D all of the profits from the innovation, patent length and breadth are not perfect substitutes, innovative knowledge may be at least partially excludable even in the absence of patent protection, etc. At a deeper level, the recent patent policy debate has made the point that the trade-off between fast innovation and static allocative inefficiency has been over-emphasized, that research on the economics of patents should focus on other issues. One big concern is that stronger patent protection may impede rather than promote technical progress when innovation is cumulative and exhibits strong complementarities (Gallini, 2002). Another major concern is that analyses based on the assumption that the patent system "works" may fail to capture the increasing complexity of modern technology and the pervasiveness of rent-seeking (Jaffe and Lerner, 2004).

In this section I develop a number of extensions of the baseline model to determine the sign, if not the magnitude, of these additional effects. Although the simple analytical framework that I use in this paper cannot do justice to the full complexity of the issues raised, the analysis may nonetheless offer some useful insight. It leads to several modified elasticity rules; for their formal derivation, see Appendix C.

5.1. Complementary innovations and patent thickets

In certain industries, such as telecommunications and biotechnology, production of new products often requires many complementary innovative components that are owned by different firms.[17] The proliferation and fragmentation of intellectual property rights creates a "patent thicket" that is often viewed as an obstacle to innovation.

Two main problems may emerge. First, a proliferation of patents held by different owners increases transaction costs and might even prevent manufacturers from obtaining the right to develop the new products, creating the tragedy of the anti-commons (Heller and Eisenberg, 1998). Second, with complementary patents there may be a problem of Cournot complements (Shapiro, 2001) that increases the deadweight loss to profit ratio D/π. How serious those problems are is a matter of controversy,[18] but it can hardly be denied that life is more difficult when a number of separate patents have to be brought together in order to operate a new technology. How do complementarity and the fragmentation of intellectual property rights impact on the optimal level of patent protection?

To the extent they are borne by firms other than the patent-holders, transaction costs call for a downward adjustment in the level of patent protection (see section 5.4 below). Instead, as discussed above (section 3.2), any rise in the deadweight loss to profit ratio impacts the optimal level of patent protection only indirectly, via the elasticity η, and does not modify the elasticity rule.

[17] For example, the IEEE 802.11 family of standards for the Wi-Fi technology involves dozens of essential patents owned by a number of separate firms (Lemley and Shapiro, 2007).

[18] Walsh *et al.* (2003) ask whether there is any indication of a tragedy of the anti-commons in biotechnology, and find evidence that transaction costs may not in fact be insurmountable. The problem of Cournot complements may be alleviated by various contractual arrangements, such as cross-licensing and patent pools (Lerner and Tirole, 2004).

Less obviously, complementarity *per se* affects the optimal level of patent protection directly and the effect is to increase that level. To see why, Box 3 considers the extreme case of strict, "two-way" complementarity – that is, each innovation has zero stand-alone value and all must be obtained in order to operate a new technology. The main conclusion is that with complementary innovations the profit ratio for *each innovator* (i.e., each one's share in the maximum hypothetical aggregate profit) should be equal to the elasticity of the supply of inventions.

This specific optimality rule rests on the strong assumptions made in Box 3, but the result that complementarity increases the optimal level of patent protection is more general. The underlying logic is that with complementary innovations, a firm investing to obtain innovation i exerts a positive externality on the firms racing to achieve innovation j. This positive externality is an additional source of distortion that tends to widen the gap between the socially optimal and the market equilibrium levels of R&D investment, making it desirable to increase the innovators' rewards as against the stand-alone case.

Box 3. Complementary innovations

Consider the case of two complementary innovations. Let x_1 denote the probability of achieving the first innovation and x_2 that of achieving the second, where the two events are statistically independent. To focus on the problem in its purest form, I assume that complementarity is strict and two-way, i.e. that each innovation has zero stand-alone value and both are indispensable to a new technology. Therefore, the new technology can be operated, and the social and private returns from it can be netted, only with probability x_1x_2. I also assume that each firm can race for either innovation 1 or innovation 2, but not both. This guarantees that intellectual property rights are fragmented.

For simplicity, I confine the analysis to the symmetric case in which the R&D expenditure function $\alpha c(x_i)$ is the same for both innovations and the private value from the innovation is split equally between the two patent-holders. With free entry in the race for each innovation, the zero-profit conditions become:

$$\tfrac{1}{2}x_i x_j v^P - \alpha c(x) = 0; \quad i,j = 1,2; \quad i \neq j$$

since a firm that achieves innovation i now obtains a positive payoff, i.e. $\tfrac{1}{2}v^P$, only if innovation j is also achieved. Note that firms racing for innovation i do not internalize the positive externality they produce for the firms racing for innovation $j \neq i$. So, there is always a no-investment equilibrium in which all firms are inactive because firms that would race for innovation i anticipate that innovation j will not be achieved, making innovation i worthless, and vice versa.

Under suitable conditions, however, there is also a stable, symmetrical, free-entry equilibrium with positive R&D investments. Assume that firms manage to coordinate on the equilibrium with positive investments if there is one – otherwise, no problem of optimal protection arises. In such an equilibrium, the zero-profit conditions reduce to

$$\tfrac{1}{2}x^2v^P - \alpha c(x) = 0$$

The social welfare function is now

$$W = x^2\left[z\frac{\pi+(1-\beta)D}{r} + (1-z)\frac{(\pi+D)}{r}\right] - 2\alpha c(x)$$
$$= v^S x^2(1-\beta z)$$

The only difference from equation (2) is that the probability that the new technology can be operated is now x^2, because two innovations have to be achieved. Appendix C shows that the optimal degree of patent protection is given by:

$$\beta z = 2\eta \tag{7}$$

i.e., *ceteris paribus* it is twice as large as in the case of the single innovation. This result extends directly to the case of n complementary innovations: the optimal strength of protection is $n\eta$, and so each patent-holder should get a fraction η of the maximum hypothetical aggregate profits. (For a free-entry equilibrium with simultaneous, positive investment in R&D to exist, a necessary condition is $n\eta < 1$, which implies that the profit ratio is necessarily lower than 1.)

With complementary innovations, there arises also a problem of appropriately dividing the aggregate profit between all the participating inventors. In the symmetric case, an even split is generally optimal, but there are many potential sources of asymmetry: R&D cost functions may differ across technological components, complementarity may not be strict, R&D investments may not be simultaneous, etc. In these cases, the optimal division of profit may be difficult to determine. Moreover, typically the policymaker does not control the division of profit precisely. As a result, certain inventors may get more than their fair share and others less.

This paper abstracts from the issue of the division of profit, but shouldn't the size of the pie and its division be determined simultaneously? Appendix C shows that when the elasticity η is constant these two issues are indeed orthogonal: the optimal aggregate profit ratio, that is to say, is given by (7) independently of the way total profits are divided. A sub-optimal split (e.g., different from fifty-fifty in the symmetric case) changes the investments in research, reducing the overall probability of success and hence social welfare, but does not affect the optimal aggregate profit ratio. (With variable elasticity things are more complicated, since changes in R&D expenditures can affect the elasticity η.)

5.2. Cumulative innovation

Perhaps the most important source of complementarity is cumulative innovation (e.g., a basic innovation plus a follow-on), when two innovations are actually needed to operate the second-generation technology. If the basic innovation has zero stand-alone value (i.e., is a pure research tool), we have precisely the case of strict, two-way complementarity analyzed above. The only difference is that now innovations must be achieved in sequence. (This is not necessarily bad, as it may facilitate firms' coordination on the equilibrium with positive investment in research: see Box 3.) Things are somewhat different if the basic innovation has a positive stand-alone value. In this case, complementarity is "one-way:" the basic innovation can be used even without the follow-on, but follow-on research cannot begin until after the basic innovation has been made.

The literature on patent policy with cumulative innovation has focused mostly on the issue of the division of profit between successive generations of innovators (see Gallini, 2002 and Scotchmer, 2004 for a thorough discussion). It is difficult, if not impossible, to ensure that all innovators have the right incentive to invest in R&D, and this literature warns that an inappropriate division of the total profit may be detrimental to technical progress. But it also suggests that cumulativeness entails a greater risk of profit erosion: unless the first-generation innovator is properly protected, competition from second-generation innovators will lower not only the former's profit but also the aggregate profit. Compounded with dynamic complementarities, which generally call for a larger total profit ratio, this effect exacerbates the under-compensation problem at the aggregate level.

There are two caveats, however. First, when innovation is cumulative the interactions between the policy tools that determine the size of the pie and those that determine its division may be subtler than in the case of static complementarity. For instance, if the first-generation patent's breadth is such that the second-generation technology infringes, forcing the two patent-holders to reach an agreement in order to exploit the second-generation technology, both the total size of the pie and the share going to the original innovator increase. Accounting for these interactions may change the optimal profit ratio with respect to the modified elasticity rule (7), but little is known on the sign and the magnitude of such changes; this is an important problem for future research.

Second, when technical progress is cumulative it may create new effects that are not easily captured by simple extensions of the baseline model. Bessen and Maskin (2007) argue, for instance, that when innovation is sequential strong patent protection reduces firms' incentive to share intermediate technological knowledge. In their model, this effect can be so strong that the rate of technical progress may be highest with no patent protection at all. Although this strong conclusion is based on various restrictive assumptions, the effect postulated may be important in certain industries.

5.3. Other protection mechanisms

Surveys conducted in various countries have repeatedly found that firms tend to rank secrecy, lead time, and the control of complementary assets ahead of patents as protection mechanisms for both product and process innovations (see e.g. Levin *et al*. 1987, Cohen *et al*. 2000).[19] This evidence is consistent with empirical estimates of the propensity to patent and the patent premium. Estimates of the propensity to patent indicate that less than half the patentable innovations are actually patented (see e.g. Duguet and Kabla, 1998). For those that are, the additional returns to innovators thanks to patent protection have been estimated to be equivalent to an implicit subsidy on R&D expenditure of 15-25% (Schankerman, 1998, McCalman, 2001, Arora *et al*. 2005). The incentive effect of patents is fairly small, and firms would certainly have substantial incentives to innovate even in the absence of patent protection. This evidently implies that the assumption that with no patent protection ($\beta z = 0$) the innovator makes zero profit is false. How does the elasticity rule change when this unrealistic assumption is relaxed?

Perhaps unexpectedly, taking other protection mechanisms into account does not necessarily lower the benchmark obtained from the baseline model. The key insight is a very simple one: with constant returns to scale, non-patent protections such as secrecy and lead time can reward innovators only to the extent that they allow pricing above marginal cost.[20] But supra-competitive pricing entails deadweight losses, irrespective of the source of market power. It follows that the other mechanisms are also distortionary, and thus their availability does not necessarily mean that society should rely less heavily on patents.

To proceed, it is convenient to distinguish between the case in which innovators can cumulate the benefits from patent protection with those of other appropriating devices and that in which other mechanisms are alternative to patent protection. In the latter case, let us focus on secrecy as an alternative to patents. (Denicolò and Franzoni, 2004 discuss the legal rules that in principle prevent innovators from relying on both patents and secrecy.) Suppose that when the innovator opts for secrecy, the invention can be concealed for a variable (and possibly uncertain) period of time. The innovator will opt to patent if and only if his expected profit under secrecy is lower than under patent protection. Assume for simplicity that the breadth of protection with secrecy is the same as with patents (but the argument is readily extended to the case where the breadth of protection differs), so that inventions for which the expected duration of the secret exceeds the patent term are kept secret and the others are patented.

[19] Cohen *et al.* (2000) report that for process innovations, only 23% of respondents consider patents an effective appropriating mechanism, as compared with 50% and 38% for secrecy and lead time, respectively. For product innovations, patents are considered relatively more effective (41%), but still less effective than either secrecy (51%) or lead time (50%).

[20] With indivisibilities or increasing marginal costs, things can be different. As is shown by Bester and Petrakis (2003) and Irmen and Hellwig (2001), if marginal costs are increasing innovators can earn positive rents even with marginal cost pricing. Boldrin and Levine (2002) argue that the same is true with indivisibilities. These theoretical results show that in some circumstances innovators may obtain a positive reward with non-distortionary means. However, it is not clear whether such non-distortionary reward is of any practical relevance, save some quite special cases.

Now, let us consider how increasing the strength of patent protection affects social welfare in this framework. By making patents relatively more attractive, such a policy measure will induce some inventors to switch from secrecy to patents. Inventions can thus be divided into three groups: those that are kept secret both before and after the measure, those that are patented both before and after, and those that switch from secrecy to patents. The patent premium is negative for innovations in the first group, positive for those in the second, and zero for "marginal" innovations in the third. A change in patent strength has no effect on the first group. For the second group, inventions that would have been patented anyway, the effect is the same as if secrecy were not a feasible option. Hence, any change in the optimal degree of patent protection relative to the baseline model must be due to the effect on the third group, marginal innovations.

More precisely, whether the secrecy option increases or decreases the optimal level of patent protection depends on whether society gains or loses when marginal innovators switch from secrecy to patent protection.

Now, if the deadweight loss from supra-competitive pricing were the only social cost associated with secrecy, for society it would be a matter of indifference whether marginal innovations are protected by patents or by secrecy, because by definition the expected duration of the secret equals the life of the patent, and the deadweight loss would be the same. In this simple case, the optimal level of patent protection is unaffected by secrecy being a feasible option.

But secrecy may entail additional social costs that make reliance on patents socially desirable. For one thing, secrecy may impede follow-on research (with patents, innovative technological knowledge must be disclosed). Second, it does not preclude independent creation by others and so invites investment to duplicate the innovation. If such duplication of efforts actually occurs, it adds to the social costs of secrecy, raising the optimal degree of patent protection above that of the baseline model.[21]

Consider next the case in which innovators can cumulate the benefits from patent with other protection mechanisms. For example, if patent-holders manage somehow to disclose innovative technological knowledge only partially in the patent specification, this may make imitation more difficult and increase effective patent breadth. Or again, brand loyalty or other entry barriers may allow patent-holders to sell their products at a premium even after the patent expires. The simplest way to take these effects into account is to re-interpret the elasticity rule (3) as a rule determining an all-inclusive profit ratio, that comprises all profits obtained by innovators through patents or other mechanisms. Of course, such a broader interpretation must also be kept in mind when assessing the current value of the profit ratio.

[21] If, however, the threat of duplication induces pre-emptive licensing of a trade secret, secrets can reward innovators more efficiently than patents (Cugno and Ottoz, 2006). But, the special difficulty of selling unprotected ideas has been recognized since Arrow (1962): see Bhattacharya and Guriev (2006) for a recent theoretical analysis. The empirical evidence seems to confirm that trade secret licensing is not very common, much less so than patent licensing: Arora and Ceccagnoli (2006), for instance, find that "the presence of a patent is almost essential for licensing" (p. 294).

5.4. Transaction costs

Intellectual property rights, like most legal institutions, entail a variety of transaction costs. The most visible costs are perhaps administrative and legal. In particular, patent litigation has recently attracted a lot of attention: it is sometimes claimed that the firms earning the most profits from patents are those with the best lawyers, not the best scientists and engineers. As a matter of fact, the number of patent lawsuits filed annually in the US doubled during the 1990s (Bessen and Meurer, 2005). The benign view of this phenomenon is that the patent litigation explosion simply reflects the upsurge in patenting (the number of lawsuits per patent seems to be roughly constant), and that in any event more aggressive enforcement deters piracy and imitation. But there is some evidence that the more a firm invests in R&D, the more likely it is to be sued (Bessen and Meurer, 2005). This may suggest that innovative firms sometimes infringe inadvertently, in which case more aggressive patent enforcement could actually discourage innovative activity. [22]

It turns out that the effect of transaction costs on the optimal level of patent protection depends on whether they are borne by patent-holders or others. Let TC denote flow transaction costs, and define $\psi = TC/(\pi + D)$ as the ratio of transaction costs to the social value of the innovation. Then, the following modified elasticity rule obtains:

$$\beta z = \frac{\eta}{1+(1-\gamma)\psi} \tag{8}$$

where γ denotes the fraction of transaction costs borne by the patent-holder. Transaction costs reduce the benchmark level of protection, but do so only to the extent that they do not fall on the patent-holder. The intuitive reason is that while all transaction costs are true social costs that reduce the value of innovations, transaction costs borne by patent-holders also reduce the private incentives to innovate.

There are various types of transaction costs, varying both in magnitude and in the extent to which they are borne by patent-holders. First, patents entail administrative costs. The application and review process is costly both for applicants and patent offices. However, to the extent that patent offices are funded by patent fees and renewal fees, most of the administrative costs are eventually borne by patent-holders and would-be patent holders. In practice, it seems that only a small fraction of the administrative costs is financed out of general revenue. These costs probably represent a very small share of the social value of innovations.

Second, patents have legal costs. Litigation is notoriously very costly. The American Intellectual Property Law Association has estimated the average cost of patent litigation at a million dollars for each party. For patents valued at more than \$25 millions, the average cost is \$4 millions for each party (Hoti *et al.*, 2006, Bessen and Meurer, 2005). These figures are certainly impressive, but considering that \$25 millions is the lower

[22] However, other interpretations are possible: for example, perhaps only firms that conduct research themselves have enough "absorptive capacity" to infringe profitably.

bound of the truncated distribution of the private values of innovations, they suggest that the ratio of transaction costs to the social value, ψ, is substantially less than one third even for patents that are litigated all the way to a decision. In the US, less than 1.5% of patents are litigated (Hoti *et al*., 2006), and even then most cases are settled before trial: less than 5% of these litigated patents go to a decision, and litigation that results in settlement is much less costly. This data strongly suggests that litigation costs are low in proportion to the value of innovations in the US, and they are unlikely to be larger elsewhere. Moreover, only half the total costs are borne by patent-holders.

Patents may entail other enforcement costs that are more difficult to measure. Most are probably associated with the prevention and detection of infringements, and as such are borne by patent-holders. Non-patentees may also sustain costs, however, e.g. to obtain technical and legal advice to avoid infringing patents, or to stand against patent-holders improperly asserting their rights.[23] In the US, there may be further costs to avoid so-called willful infringement. Finally, there are transaction costs associated with patent licensing, part of which must be borne by licensees. These licensing costs can be large, especially if intellectual property rights are highly fragmented. To my knowledge, however, there is no quantitative evidence.

5.5. Business stealing and other patent pathologies

The baseline model posits that the patent-holder's profit is a share of what he contributed to society. This implicitly assumes that he can only exploit users of the new product or device and cannot make non-users worse off. This assumption may fail to hold for various reasons, such as business stealing, the enforcement of dubious patents, and hold-up.

5.5.1. Business stealing

If in the pre-innovation equilibrium the industry comprises incumbents holding some market power, the innovator may be able to steal at least part of their rents. For example, the author of a new economics textbook may subtract some royalties from authors of competing textbooks (although he will also profit from the increase in the buyers' average willingness to pay associated with the enlarged variety of available products). This is the standard business stealing effect.

5.5.2. Dubious patents

Jaffe and Lerner (2004) argue that seemingly innocuous reforms – like the institution in 1982 of new Courts of Appeals for the Federal Circuit, creating a unified appellate authority for all patent cases and the reform of US Patent Office (USPTO) funding system in the early 1990s – substantially weakened the standards of novelty and non-

[23] Bessen and Meurer (2007) use a large set of event studies to estimate the expected loss in market value that firms suffer when they are sued for patent infringement. The mean loss is large, almost $30 millions, but it is unclear whether these losses are due to transaction costs, or they simply reflect the risk that the lawsuit terminates the profits accruing to the firm accused of infringement and reduces the value of its specific investments.

obviousness for patentability in the US. The USPTO is issuing many dubious patents, [24] so the argument goes, and the likelihood of their being upheld by the courts has increased.[25] This leaves abundant scope for opportunistic behaviour by firms that seek patent protection even when they have not achieved genuine innovations.

The benign view here is that patents that are clearly invalid, of which examples undeniably abound, are unenforceable and so confer no effective market power. Lemley (2001) argues further that in-depth screening of all patent applications would be inefficient. Since the vast majority of patents is valueless and would never be enforced anyway, it is more efficient to use scarce resources for judicial review of patents that are enforced and litigated.

Farrell and Shapiro (2005) counter that even "weak" patents (i.e., patents that have a low probability of surviving judicial scrutiny) can command substantial market power for two reasons: free riding and collusion (see also Chiou, 2006). Free riding may arise because filing a suit is costly, while all potential users of the technology benefit from a patent being found invalid. Collusion can also be a problem, because both parties may have an interest in a settlement rather than a win-or-lose decision, and because even weak patents can be licensed strategically to support anticompetitive outcomes. Thus, weak patents might give opportunistic agents the power to tax consumers and possibly also genuine inventors, whose incentives to innovate would be correspondingly reduced.

5.5.3. Hold-up

A related problem is hold-up. The design of innovative products often exhibits a putty-clay pattern: *ex ante*, the product can be easily designed in various alternative ways, but it may be difficult and time-consuming to modify the design *ex post*. With complex technologies characterized by fragmented intellectual property rights, manufacturers may inadvertently infringe some patents that they could easily have circumvented in advance (Shapiro, 2006). After the fact, however, patent-holders whose property rights have been inadvertently violated may acquire a lot of bargaining power, especially if they can enjoin manufacturers from marketing the products at issue. As a result, patent-holders can extract disproportionately large royalties with respect to the real value of their innovations, obtaining rents that otherwise would accrue to somebody else.

The prevalence of the hold-up problem is controversial. Some commentators talk of legions of opportunistic patent-holders – so called patent trolls – that purportedly delay the filing of patent lawsuits so as to take advantage of the strong bargaining power they may acquire once manufacturers have sunk substantial investments. But the available evidence is exclusively anecdotal and it is very hard to gauge the importance of this phenomenon accurately.

---

[24] In principle the patent system may also err in the opposite direction, i.e., by not granting a patent for genuine inventions. The probability of a genuine invention's being denied protection should also be taken into account when assessing patent breadth, but currently it would appear to be negligible.

[25] It must be said that Europe seems relatively immune from these purported defects of the American patent system. The patents granted by the European Patent Office seem to be less dubious than those allowed in the US, although the European Patent Office too suffers from overload and lack of appropriate incentives. Moreover, the European post-grant opposition procedure is effective in pruning low-quality patents (Friebel *et al.*, 2006). Finally, the decline in the quality of US patents seems, at least in part, to be due to the expansion in patentable subject matter, which has been less pronounced in Europe.

5.5.4. Analysis

To analyze the impact of business stealing, dubious patents, and hold-up on the optimal level of patent protection, assume that only a fraction 1-ϕ of the innovator's flow profit π corresponds to net social value. The parameter ϕ can be thought of as the coefficient of business stealing and hold-up, i.e., the fraction of the innovator's rents that are taken from incumbents or manufacturers. Alternatively, it can be interpreted as the share of bad patents, i.e., the probability of a patent's being granted and enforced in the absence of a genuine innovation. The elasticity rule then becomes:

$$\beta z = \frac{(1-\phi)\pi + D}{\pi + D}\eta \tag{9}$$

The benchmark level of protection is now lower than in the baseline model.

An upper bound on the required adjustment can be obtained by setting $D = 0$ in the modified elasticity rule (9). In this case, the optimal profit ratio is $(1 - \phi)$ times the elasticity of the supply of inventions. Unfortunately, the evidence is mainly anecdotal and so the magnitude of the parameter ϕ is a guess at best.

5.6. The optimal combination of breadth and length

In the baseline model, the deadweight loss is proportional to the breadth of patent protection. While a positive correlation between the two variables is natural, the proportionality assumption is restrictive. For example, in Gilbert and Shapiro (1990) the deadweight loss is a convex function of patent breadth, and in Gallini (1992) it is concave.

To cover these cases, assume that the deadweight loss while the patent is in force is an increasing function of the breadth of protection, $D(\beta)$, with $D(1) = D$ and $D(0) = 0$. The baseline model assumes $D(\beta) = \beta D$, which implies that the combination of breadth and length is indifferent. When the function $D(\beta)$ is not affine, a non-trivial problem of choosing the optimal combination of breadth and length arises. Typically, this problem has a corner solution with either $\beta = 1$ or $z = 1$ (see Denicolò, 1996 for a survey).[26] Which is optimal depends on the sign of the second derivative $D''(\beta)$. If the function $D(\beta)$ is concave, the policymaker sets $\beta = 1$. Social welfare then reduces to $W = v^S x(1 - z)$ and the elasticity rules (3) applies with no changes. If instead the function $D(\beta)$ is convex, the policymaker sets $z = 1$. In this case there is no closed-form solution for the optimal breadth, but Appendix C shows that the optimal profit ratio is higher than in the baseline model.

[26] This is clearly unsatisfactory, since in real life both patent breadth and length are limited. So far, theoretical models of the optimal combination of breadth and length have not been able to generate meaningful interior optima that mirror observed policy choices; as a result, they have had little impact on policy.

5.7. Monopoly in research

The baseline model assumes that there is free entry in research, which means that all profits from the innovation are invested in R&D. This assumption is probably overoptimistic. If the share of profits invested in research is lower than one, e.g. because competition in research is limited, innovative firms will obtain positive net rents. However, it is clear that the elasticity rule (3) would continue to hold if such a share were constant and the policymaker maximized the expected surplus for consumers.

With monopoly in research, two complications arise. First, the monopolist's profits may be included in the social welfare function, as is implicit in equation (2). In this case, the optimal level of patent protection with monopoly is higher than in the baseline model because now expected profits are not entirely dissipated in the patent race. Second, the share of expected profits invested in research – which now coincides with the elasticity η – need not be constant. Generally speaking, the sign of this latter effect is uncertain.

Because of these new effects, with monopoly in research the elasticity rule becomes:

$$\beta z = \frac{\pi + D}{\eta\pi + D - \varepsilon\eta D}\eta \qquad (10)$$

where $\varepsilon = (d\eta/dx)(x/\eta)$ is the elasticity of η with respect to the probability of success. Note first of all that with $D = 0$ the social problem now has a corner solution $\beta z = 1$; the intuitive reason for this is explained above (footnote 10).

With an iso-elastic R&D cost function ($\varepsilon = 0$), the share of profits invested in research is constant. In this case the optimal profit ratio under monopoly is higher than under free entry in research, because R&D costs cover only a share η of the expected profits from the innovation, and so the remaining share is net social welfare. Formally, this effect is captured by the term $\eta < 1$ that multiplies π in the denominator of (10).

If $\varepsilon > 0$, the share of profits re-invested in research increases with the level of R&D investment; as a consequence, the stimulus to innovative activity provided by patent protection is more powerful and the optimal level of protection increases. In this case, both of the additional effects discussed above raise the optimal profit ratio. Moreover, even if $\varepsilon < 0$ the optimal strength of protection under monopoly can be greater than under free entry, as long as the elasticity ε is not too low – to be precise, the condition is $\varepsilon > (\eta - 1)\pi/\eta D$.

Although the case of monopoly in research is extreme, the effects highlighted above are more general. Using the Poisson model (which implies $\varepsilon < 0$), I have shown elsewhere that in intermediate cases between monopoly and free entry, the weaker competition in the invention industry is, the higher is the optimal profit ratio (Denicolò, 1999).

5.8. Risk premium

Another reason why firms may not invest in research all of the expected profits from innovation is risk. Since innovative activity is risky and capital markets are imperfect, the rate at which firms discount future profits, which is denoted here by ρ, may include a substantial risk premium. Thus, the social and private discount rates may differ, with $\rho > r$. The analysis of optimal patent protection when the two rates differ is generally complex, but some insight can be gained by looking at the simple case in which $T = \infty$.

With $T = \infty$, the profit ratio coincides with patent breadth, β, and the free-entry condition implies that R&D expenditure is equal to $x\beta\pi/\rho$. Thus, firms re-invest in research a constant share $r/\rho < 1$ of the "true" expected profits from the innovation, $x\beta\pi/r$. By the same logic as in the case of research monopoly, the optimal profit ratio is

$$\beta = \frac{\pi + D}{\frac{r}{\rho}\pi + D}\eta \tag{11}$$

Clearly, the optimal profit ratio is higher than in the baseline model. This has an intuitive explanation: firms now re-invest in research only a share of the "true" expected profits from the innovation and the remaining share is net social welfare.

However, with a risk premium patent breadth and length are no longer perfect substitutes and so setting $T = \infty$ may not be an innocuous approximation. A more complete analysis of the effects of risk is left for future work.

5.9. R&D subsidies

Many governments subsidize R&D expenditure. For instance, according to Guellec and Van Pottelsberghe De La Potterie (2003) the share of business-performed R&D that is financed by the government through direct subsidies or preferential tax treatment is around 25% in the US. With R&D subsidies, the modified elasticity rule is:

$$\beta z = \frac{\pi + D}{\frac{1}{1-s}\pi + D + \frac{\lambda s}{1-s}\pi}\eta \tag{12}$$

where s is the subsidy rate and λ is the excess burden of taxation per unit of fiscal revenue (thus, $1 + \lambda$ is the shadow cost of public funds). Clearly, the optimal profit ratio is lower than in the baseline model.

There are two negative effects of R&D subsidies on the optimal level of protection. The first effect is that firms now re-invest in R&D a multiple of the expected profits from the innovation because they sustain only a share $(1 - s)$ of the R&D costs. This effect, which operates even if $\lambda = 0$, is captured by the term $1/(1 - s) > 1$ that multiplies π in the denominator of (12). The effect is similar in nature, but opposite in sign, to the effect of risk and research monopoly discussed above. Generally speaking, the optimal profit ratio is lower than in the baseline model if firms re-invest in R&D more than the expected profits from the innovation, higher if they re-invest less. Which is more likely? I am inclined to believe that firms re-invest less, but more empirical evidence is needed to assess the issue.

The second effect arises when R&D subsidies are financed through distortionary taxes: $\lambda > 0$. In this case, other things being equal, stronger patent protection increases R&D spending and hence subsidies, which increases the excess burden of taxation. This additional social cost of patent protection lowers the optimal profit ratio.

An upper bound on this latter effect can be obtained by setting $D = 0$ in the modified elasticity rule (12). In this case, neglecting the former effect discussed above, the extra term in the denominator is just $\lambda s/(1 - s)$. The parameter λ is typically estimated in the range 20-30%. With $s = 0.25$ and $\lambda = 0.25$, the extra term is 1/12; thus, the optimal profit ratio is *at least* 12/13 of the elasticity of the supply of inventions. These back-of-the-envelop calculations suggest that the adjustment required to account for distortionary taxation is on the order of 10% or less.

5.10. Summing up

In this section I have argued that with cumulative or complementary innovations the profit ratio should be greater than in the baseline model because of the positive externality that the firms racing for an innovation generate for those racing for complementary or subsequent innovations. Nor does availability of other mechanisms, such as secrecy or lead time, to appropriate the returns from investment in research mean that patent protection should be weakened, as long as those mechanisms too are distortionary. Transaction costs and R&D subsidies unambiguously lower the optimal level of patent protection. However, only the share of transaction costs sustained by firms other that the patent holder calls for a downward adjustment in the profit ratio. Moreover, most of the effect of R&D subsidies is probably offset by other countervailing effects, like risk and imperfect competition in research. If this is true, then the magnitude of the required adjustment is likely to be small. Business stealing and opportunistic behaviour by firms seeking patent protection also call for a downward adjustment in the profit ratio, the magnitude of which is however unknown. Overall, the analysis suggests that the baseline model is not severely biased against the over-reward hypothesis.

6. CONCLUSION

Many economists believe that a market economy undertakes too little research. According to Jones and Williams (1998), if it were possible to increase R&D investment by fiat, it should be at least doubled. Unfortunately, we live in a second-best world where this is not possible and, as a result, even highly imperfect policy tools like patents can then be useful. This paper asks whether in such a world we haven't gone too far in using patents to incentivize innovative activity; that is, whether patents over-compensate innovators.

To address this issue, I have developed a simple but flexible model of the optimal level of patent protection that captures the basic trade-off between rapid innovation and monopoly distortions. The baseline model implies that innovators are over-compensated if the profit ratio exceeds the elasticity of the supply of inventions. The innovation production function literature suggests that a reasonable range for the elasticity of the supply of inventions is from 0.5 to 0.7. On the other hand, it seems unlikely that the profit ratio will exceed 0.5 in current practice, and even 0.3 might be considered a somewhat optimistic guess. This strongly suggests that the over-compensation hypothesis is not supported by the data.

Moving beyond the baseline model, many other determinants come into play. In some extensions, the optimal level of patent protection turns out to be higher than in the baseline model, reinforcing the sensation gathered from baseline analysis even in the absence of precise quantitative estimates of the additional parameters. Other effects are ambiguous in sign or negative, but unfortunately little is known on their magnitude. More empirical research is therefore needed for a more precise assessment, but the analysis suggests that the preliminary assessment is not severely biased against the over-reward hypothesis.

Overall, a preponderance of what evidence is currently available points against the over-reward hypothesis. It is tempting to conclude that policy reform, if anything, should strengthen patent protection. At this stage, however, no policy conclusion can be anything but tentative. The assessment developed in this paper, while highly suggestive, is not truly compelling: reasonable interpretations of the same evidence (or lack thereof) might differ. Moreover, the analysis is based on a very simple model that cannot capture such important and complex issues as the division of profit among different innovators, the sharing of intermediate technological knowledge, the strategy of amassing large patent portfolios for defensive reasons (Hall and Ziedonis, 2001), the different incentives of incumbents and laggards to innovate, and the related possibility that the optimal level of patent protection could depend on the size of the leader's competitive advantage (Acemoglu and Akcigit, 2006). More theoretical work is needed to understand how these and other subtle effects impact on the optimal level of patent protection.

In spite of these limits, this paper shows that useful insights can be obtained by linking theoretical models of optimal patent design to the empirical literature. Even if future research were to overturn this paper's specific conclusions, the simplicity and flexibility of the elasticity rule might well persuade other researchers that a more accurate assessment of the over- or under-reward hypotheses is not beyond reach. Its application could help clarify what additional empirical research is most needed to provide reliable answers.

Appendix A: Alternative models of innovation

In this Appendix I show that equation (1) can be regarded as a reduced form of various seemingly different models of innovation that are commonly employed in the literature.

A1. The Poisson model

Consider first the standard patent race model where the timing of the innovation is a probabilistic function of the amount invested in R&D by research firms (Loury, 1979; Dasgupta and Stiglitz, 1980). At the beginning of the patent race, each participating firm i decides its R&D effort y_i and pays a lump sum cost αy_i, where α is the constant marginal cost of R&D effort. The R&D effort determines the expected time of successful completion of the R&D project according to a Poisson discovery process. Thus, the payoff function of firm i (i.e., the present value of expected profits, net of R&D costs) is:

$$\begin{aligned}\Pi_i &= \int_0^\infty e^{-(Y+r)t} y_i v^P dt - \alpha y_i \\ &= \frac{y_i v^P}{Y+r} - \alpha y_i\end{aligned} \tag{A.1}$$

where y_i is i's R&D effort and also i's instantaneous probability of innovating, conditional on no previous success, $Y = \sum y_i$ is the aggregate R&D effort, and e^{-Yt} is the probability that no firm has yet innovated by time t.

Aggregate profits are $Yv^P/(Y + r) - \alpha Y$. Here $x = Y/(Y + r)$ can be thought of as the "discounting adjusted" probability of success: with a Poisson discovery process the innovation eventually occurs with probability one, but since there is discounting, a delayed success is valued less than instant success. The associated R&D cost function $c(x)$ is implicitly defined by the equation $Y = rx/(1 - x)$. The function $c(x)$ is increasing, convex, and satisfies $c(0) = 0$. Thus, aggregate profits can be rewritten as $xv^P - \alpha c(x)$.

With free entry in the R&D industry, the zero-profit condition (1) then determines the aggregate R&D effort $Y = \max\,[v^P/a - r, 0]$ or, equivalently, $x = \max\,[1 - r\alpha/v^P, 0]$. With constant returns in research, the equilibrium number of active firms and their respective R&D investment are indeterminate, and only aggregate R&D investment is determined. However, this model readily extends to the case of variable returns, monopoly in research, and competition with a fixed number of research firms (see Denicolò, 1999 for details).

A2. Variable timing

Assume that the date of innovation t is a deterministic function of the R&D investment. The innovator's profits are $v^P e^{-rt} - \alpha\hat{c}(t)$, where $\alpha\hat{c}(t)$ is the discounted cost of innovating at time t. Assume that $\lim_{t\to 0} \hat{c}(t) = \infty$ and that $\hat{c}(t)$ decreases rapidly enough with t. Define $x = e^{-rt}$, such that $t = -(\log x)/r$. The profit function then becomes $xv^P - \alpha c(x)$, where $c(x) = \hat{c}[-(\log x)/r]$. If there is free entry in research and a patent is granted to the first firm that innovates, the date of innovation will be chosen such that the discounted profit $xv^P - \alpha c(x)$ vanishes. With monopoly in research, the timing of the innovation will be chosen so as to maximize $xv^P - \alpha c(x)$.

A3. Fixed, uncertain R&D costs

Here I assume there is monopoly in research (see Appendix C below). Following Scotchmer (2004), suppose that there is an R&D project that can be undertaken at a cost αc and generates instantaneously and for sure an innovation of private value v^P. The cost c is a random variable drawn from a distribution $G(c)$ with support $[0, c]$ and density $g(c)$. The realization of c is known to the firm prior to the investment decision. Clearly, the firm invests when the realization of the cost is below a cut-off value. Denote this cut-off by $G^{-1}(x)$, so that x is the *ex ante* probability that the innovation is achieved. The expected R&D expenditure is then $c(x) = \int cg(c)\mathrm{d}c$, where the integral is taken over the interval $[0, G^{-1}(x)]$.

In particular, the firm will invest if and only if $\alpha c \leq v^P$, and so the *ex ante* probability of success is $G(v^P/\alpha)$. But this is equivalent to assuming that the research firm maximizes its expected profits $xv^P - \alpha c(x)$. To see this, it suffices to note that the first-order condition for this maximization problem is $v^P = \alpha(\mathrm{d}G^{-1}(x)/\mathrm{d}x)G^{-1}(x)g[G^{-1}(x)]$, or equivalently $v^P = \alpha G^{-1}(x)$. Thus, the optimal cut-off is indeed v^P/α.

A4. Variable quality

As a final example, consider a firm that can create a new product of variable quality. The variable x now denotes the quality of the new product, and π denotes the profit per unit of quality. If $\alpha c(x)$ is the cost of creating a good of quality x, the firm's profit function will be $xv^P - \alpha c(x)$. With monopoly in research, the firm will simply choose the quality level that maximizes its profits. (Stretching the interpretation of the model, one could imagine that if a patent is granted to the firm that achieves the highest quality level, with free entry in research a quality level will be chosen such that the profit is driven to zero.)

Appendix B: The elasticity rule

First of all, note that both x and W depend on β and z only through the profit ratio βz. Differentiating (2) with respect to βz one gets the following first-order condition for a maximum:

$$\frac{dW}{d(\beta z)} = \frac{\pi + D}{r}\left[\frac{dx}{d(\beta z)}(1-\beta z) - x\right] = 0 \tag{B1}$$

Using the zero-profit condition (1), by the implicit function theorem one gets

$$\frac{dx}{d(\beta z)} = \frac{\eta}{1-\eta}\frac{x}{\beta z} \tag{B2}$$

Substituting into (B1) and rearranging one obtains

$$\left[\frac{\eta}{1-\eta}\frac{1-\beta z}{\beta z}-1\right]=0 \tag{B3}$$

whence equation (3) follows immediately.

B1. Heterogeneous innovations

Here I extend the elasticity rule (3) to the case of heterogeneous innovations. Let innovations differ according to a single parameter σ that varies on a domain Σ. (The case of multi-dimensional heterogeneity is similar, with multiple integrals replacing simple integrals whenever appropriate.) Let $\pi(\sigma)$, $D(\sigma)$ and $c(x, \sigma)$ denote the flow profit, deadweight loss and R&D cost function for an innovation of type σ. Let $F(\sigma)$ be the distribution function of innovations type and let $f(\sigma)$ be the associated density function. For each innovation σ, a zero-profit condition like (1) must hold, i.e.,

$$x\beta z\frac{\pi(\sigma)}{r}-\alpha c(x,\sigma)=0 \tag{B4}$$

These conditions determine the equilibrium level of investment in research as a function of innovation type σ. The social welfare function becomes

$$\begin{aligned} W &= \int\left[xz\frac{\pi(\sigma)+(1-\beta)D(\sigma)}{r}+x(1-z)\frac{\pi(\sigma)+D(\sigma)}{r}-\alpha c(x,\sigma)\right]f(\sigma)d\sigma \\ &= \int\frac{\pi(\sigma)+D(\sigma)}{r}x(\sigma)(1-\beta z)f(\sigma)d\sigma \end{aligned} \tag{B5}$$

where integrals are taken over the domain Σ. Note that (B2) holds for each innovation, and so differentiating (B5) one gets the following generalization of (B3):

$$\frac{dW}{d(\beta z)}=\int\ \frac{\pi(\sigma)+D(\sigma)}{r}x(\sigma)\left[\frac{\eta(\sigma)}{1-\eta(\sigma)}\frac{1-\beta z}{\beta z}-1\right]f(\sigma)d\sigma=0 \tag{B6}$$

whence it is clear that if the elasticity $\eta(\sigma)$ is constant across innovations, one re-obtains the elasticity rule (3). More generally, rearranging terms one now obtains:

$$\beta z=\frac{\int\frac{[\pi(\sigma)+D(\sigma)]x(\sigma)}{1-\eta(\sigma)}\eta(\sigma)f(\sigma)d\sigma}{\int\frac{[\pi(\sigma)+D(\sigma)]x(\sigma)}{1-\eta(\sigma)}f(\sigma)d\sigma} \tag{B7}$$

The right hand side of (B7) can be interpreted as a weighted average of the individual elasticities $\eta(\sigma)$, with weights reflecting the value of innovations and their relative frequency. The weights are increasing in $\eta(\sigma)$, implying that a mean preserving spread in the distribution of the elasticities raises the optimal level of protection.

B2. The private and social rates of return to R&D

Next, I analyze the relationship between the profit ratio and the private and social rates of return to R&D, showing that the elasticity rule implies a lower bound for the ratio between the two.

To begin with, I relate the private and social rates of returns to RD to the variables of the theoretical model. First, note that with free-entry, firms conducting research earn no extra-profits and so the private rate of return to R&D is r. The social rate of return, r_S, is the discount rate that equates the social benefits from innovation to the R&D costs. Thus, r_S is implicitly defined as the solution to the following equation:

$$\frac{\pi + CS + (1-\beta z_S)D}{r_S} - \alpha c(x) = 0 \tag{B8}$$

where z_S is the normalized length associated with r_S and I now allow for $CS > 0$. From (1) and (B8) it follows that

$$\frac{r}{r_S} = \frac{\pi}{\pi + CS + (1-\beta z_S)D}\beta z \tag{B9}$$

This means that the ratio between the private and social returns to R&D is actually lower than the profit ratio. The intuitive reason is that technological spillovers, consumer surplus and other positive externalities associated with innovative activity would create a gap between the social and private returns to R&D even if the profit ratio were equal to 1.

To calculate a proper benchmark for the ratio between the private and social rates of returns, however, one cannot simply substitute (B9) into the elasticity rule (3), for the positive externalities also affect the optimal level of patent protection. To proceed, I therefore now derive the modified elasticity rule for the case $CS > 0$. With $CS > 0$, the social welfare function becomes

$$\begin{aligned} W &= x\left[z\frac{\pi + (1-\beta)D + CS}{r} + (1-z)\frac{(\pi + D + CS)}{r}\right] - \alpha c(x) \\ &= \frac{(\pi + D)}{r}x(1-\beta z) + x\frac{CS}{r} \end{aligned} \tag{B10}$$

Using (B2), the first-order condition for a maximum becomes

$$\frac{\pi + D}{r}\left[\frac{\eta}{1-\eta}(1-\beta z + \sigma)\frac{x}{\beta z} - x\right] = 0 \tag{B11}$$

whence one gets:

$$\beta z = \frac{\pi + CS + D}{\pi + D}\eta \tag{B12}$$

The profit ratio should now be higher than the elasticity of the supply of inventions. This result makes intuitive sense: any positive externality arising from innovative activity calls for stronger incentives to invest in research.

Combining the modified elasticity rule (B12) and (B9), one finally obtains:

$$\frac{r}{r_S} = \frac{\pi}{\pi + D}\frac{\pi + CS + D}{\pi + CS + (1 - \beta z_S)D}\eta \tag{B13}$$

This equation provides an appropriate benchmark for the ratio between the private and social returns to R&D. As discussed above, this benchmark is necessarily lower than 1 (footnote 1). From (B13), inequality (5) in the text immediately follows.

Appendix C: Modified elasticity rules

C1. Complementary innovations

Using the zero-profit conditions derived in Box 3, i.e., $\frac{1}{2}x^2\beta z\pi/r - \alpha c(x) = 0$, one now obtains:

$$\frac{dx}{d(\beta z)} = \frac{\eta}{1 - 2\eta}\frac{x}{\beta z} \tag{C1}$$

The social welfare function has already been derived in Box 3. Differentiating and using (C1), the modified elasticity rule (7) immediately follows.

To analyze the case of an arbitrary division of the total profit π between the two patent-holders, suppose that the first gets a share κ of π and the second the remainder, $(1 - \kappa)$. Assume that the R&D expenditure function $c(x)$ is iso-elastic: $c(x) = x^h$, where $h = 1/\eta$. The two zero-profit conditions become $\kappa x_1 x_2 \beta z\pi/r - \alpha_1 c(x_1) = 0$ and $(1 - \kappa)x_1 x_2 \beta z\pi/r - \alpha_2 c(x_2) = 0$, respectively. From these equations one gets

$$x_1 = \left[\frac{\kappa\alpha_2}{(1 - \kappa)\alpha_1}\right]^\eta x_2 \tag{C2}$$

which can be substituted back into the zero-profit conditions to get

$$(1 - \kappa)\left[\frac{\kappa\alpha_2}{(1 - \kappa)\alpha_1}\right]^\eta v^P = \alpha_2 x_2^{h-2} \tag{C3}$$

Solving for x_1 and x_2 and substituting into the social welfare function one gets

$$W = \frac{(\pi + D)}{r}\left[\frac{\kappa(1-\kappa)}{\alpha_1\alpha_2}\right]^{\frac{\eta}{1-2\eta}}\left(\frac{\beta z\pi}{r}\right)^{\frac{2\eta}{1-2\eta}}(1-\beta z) \tag{C4}$$

whence one obtains again the modified elasticity rule (7), independently of κ. This proves that in this extension the issue of the division of profit is indeed orthogonal to the issue of the optimal aggregate level of protection.

C2. Transaction costs

For simplicity, assume that transaction costs are proportional to the strength of patent protection, so that a fraction β of transaction costs TC is borne until the patent lapses. Assume also that the patent-holder sustains a fraction γ of total transaction costs. Then, the zero-profit condition becomes

$$x(v^P - \gamma\beta zTC/r) - \alpha c(x) = 0 \tag{C5}$$

The social welfare function becomes

$$\begin{aligned} W &= x\left[z\frac{\pi + (1-\beta)D - \beta TC}{r} + (1-z)\frac{\pi + D}{r}\right] - \alpha c(x) \\ &= x\left[\frac{\pi + D}{r}(1-\beta z) - \beta z\frac{(1-\gamma)TC}{r}\right] \end{aligned} \tag{C6}$$

where the second line of (C6) follows by inserting the zero-profit condition (C5) into the first line. The first-order condition for a maximum becomes

$$\frac{\pi + D}{r}\left[\frac{dx}{d(\beta z)}(1-\beta z - \beta z(1-\gamma)\psi) - x - x(1-\gamma)\psi\right] = 0 \tag{C7}$$

By implicit differentiation of (C5), it can be shown that equation (B2) continues to hold. From (C7), using (B2) one obtains

$$\frac{\pi + D}{r}\left[\frac{\eta}{1-\eta}(1-\beta z - \beta z(1-\gamma)\psi)\frac{x}{\beta z} - x - x(1-\gamma)\psi\right] = 0 \tag{C8}$$

Rearranging and simplifying, the modified elasticity rule (8) easily follows.

C3. Business stealing

With business stealing, dubious patents, and hold-up, a share ϕ of the patent-holder's profits is a pure transfer and does not represent net social welfare; as such, it is not included in the social welfare function. The social welfare function then becomes

$$W = x\left[z\frac{(1-\phi)\pi + (1-\beta)D}{r} + (1-z)\frac{(1-\phi)\pi + D}{r}\right] - \alpha c(x) \qquad \text{(C9)}$$

Using the zero-profit condition (1), the social welfare function (C9) now simplifies to

$$W = \frac{(\pi + D)}{r}x(1-\beta z) - x\phi\frac{\pi}{r} \qquad \text{(C10)}$$

Since the zero-profit condition (1) does not change, (B2) continues to hold. Then, the first-order condition for a maximum becomes

$$\frac{\pi + D}{r}\left[\frac{\eta}{1-\eta}\left(1-\beta z - \phi\frac{\pi}{\pi + D}\right)\frac{x}{\beta z} - x\right] = 0 \qquad \text{(C11)}$$

Simple algebra then suffices to obtain the modified elasticity rule (9).

It must be said that the social welfare function (C9) abstracts from some potentially important effects of business stealing, dubious patents and hold-up. First, it does not take into account that business stealing is usually associated with more intense competition in the product market, which, other things being equal, may reduce the profit ratio and may also lower the deadweight loss as against the baseline model. Second, to the extent that dubious patents and hold-up may allow patent-holders to tax genuine innovators, they may negatively affect the incentives to innovate. The analysis developed here implicitly assumes that these effects are negligible or almost exactly cancel.

C4. The optimal combination of breadth and length

Following the discussion in the text, assume that the deadweight loss is an increasing function of the breadth of protection, $D(\beta)$, with $D(1) = D$ and $D(0) = 0$. The proof that $D''(\beta) < 0$ (respectively, $D''(\beta) > 0$) entails $\beta = 1$ (respectively, $z = 1$) can be found in Denicolò (1996) and will not be repeated here. Rather, I focus on the case of $D''(\beta) > 0$ and so I optimally set $z = 1$. Using the zero-profit condition, social welfare then reduces to

$$W = x\left[\frac{(1-\beta)\pi}{r} + \frac{D - D(\beta)}{r}\right] \qquad \text{(C12)}$$

The first-order condition for a maximum becomes

$$\frac{dx}{d\beta}\left[\frac{(1-\beta)\pi}{r} + \frac{D - D(\beta)}{r}\right] - x\frac{\pi + D'(\beta)}{r} = 0 \qquad \text{(C13)}$$

Simplifying and rearranging one gets the following modified elasticity rule

$$\beta = \frac{\eta}{\Xi + \eta(1-\Xi)} \qquad \text{(C14)}$$

where

$$\Xi := \frac{\pi + D'(\beta)}{\pi + \dfrac{D - D(\beta)}{1-\beta}} \tag{C15}$$

Although (C14) does not yield a closed-form solution for β, note that $D''(\beta) > 0$ implies $D'(\beta) < [D - D(\beta)]/(1 - \beta)$. It follows that $\Xi < 1$ and so $\beta > \eta$. Since $z = 1$, this means that the optimal profit ratio is now greater than the elasticity of the supply of inventions.

C5. Monopoly in research

With monopoly in research, the monopolist maximizes $xv^P - \alpha c(x)$. Thus, the zero-profit condition $xv^P - \alpha c(x) = 0$ is replaced by the first-order condition:

$$v^P - \alpha c'(x) = 0 \tag{C16}$$

The social welfare function is

$$W = (1/r)x[\pi + (1 - \beta z)D] - \alpha c(x) \tag{C17}$$

The first-order condition for maximizing W under the constraint (C16) is

$$\frac{c'(x)}{xc''(x)}\frac{1-\beta z}{\beta z} = \frac{D}{\pi + D} \tag{C18}$$

It can be shown that

$$\frac{xc''(x)}{c'(x)} = \frac{1}{\eta} - 1 - \varepsilon \tag{C19}$$

where $\varepsilon = (d\eta/dx)(x/\eta)$. Using (C19), rearranging (C18) and simplifying one gets the modified elasticity rule (10).

C6. R&D subsidies

If firms conducting R&D are subsidized at rate s, the zero-profit condition becomes:

$$xv^P - (1 - s)\alpha c(x) = 0 \tag{C20}$$

Taking into account the excess burden of taxation, the social welfare function becomes

$$\begin{aligned} W &= x\left[z\frac{\pi + (1-\beta)D}{r} + (1-z)\frac{\pi + D}{r}\right] - \alpha c(x) - \lambda s \alpha c(x) \\ &= \frac{x}{r(1-s)}\left[(1 - s - \beta z - \lambda\beta z s)\pi + (1-s)(1-\beta z)D\right] \end{aligned} \tag{C21}$$

Using (B2), the first order condition for a maximum becomes

$$\frac{\eta}{1-\eta}\frac{(1-\beta z)(\pi+D)-s(1+\lambda\beta z)\pi-s(1-\beta z)D}{\beta z}-(\pi+D+\lambda s\pi+sD)=0 \tag{C22}$$

Rearranging and simplifying one obtains the modified elasticity rule (12).

REFERENCES

Acemoglu, D. and U. Akcigit (2006). 'State-dependent intellectual property rights policy', mimeo, MIT.

Acemoglu, D. and J. Linn (2004). 'Market size in innovation: Theory and evidence from the pharmaceutical industry', *Quarterly Journal of Economics*, 119(3), 1049-1090.

Agarwal, R. and M. Gort (2001) 'First-mover advantage and the speed of competitive entry, 1887-1986', *Journal of Law and Economics*, 44(1), 161-177.

Arora, A., Ceccagnoli M. and W. M. Cohen (2005). 'R&D and the patent premium', W. P. No. 9431, National Bureau of Economic Research, revised 2005.

Arora, A. and M. Ceccagnoli (2006). 'Patent protection, complementary assets, and firms' incentives for technology licensing', *Management Science*, 52(2), 293-308

Arrow, K. (1962). 'Economic welfare and the allocation of resources for invention', in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by R. R. Nelson, Princeton, N.J., Princeton University Press, 609-625.

Baumol, W. (2002). *The free market innovation machine*, Princeton, N.J., Princeton University Press.

Bessen, J. and R. Hunt (2007). 'An empirical look at software patents', *Journal of Economics and Management Strategy*, 16(1), 157-189.

Bessen, J. and E. Maskin (2007). 'Sequential innovations, patents, and imitation', forthcoming in *RAND Journal of Economics*.

Bessen, J. and M. J. Meurer (2005). 'The patent litigation explosion', mimeo, Boston University School of Law.

Bessen, J. and M. J. Meurer (2007). 'The private costs of patent litigation', mimeo, Boston University School of Law.

Bester, H. and E. Petrakis, (2003). 'Wages and productivity growth in a competitive industry', *Journal of Economic Theory*, 109, 52-69.

Bhattacharya, S. and S. Guriev (2006). 'Patents vs. trade secrets. Knowledge licensing and spillovers', *Journal of European Economic Association,* 4(6), 1112-1147.

Blundell, R., Griffith R. and F. Windmeijer (2002). 'Individual effects and dynamics in count data models', *Journal of Econometrics*, 108(1), 113-131.

Boldrin, M. and D. Levine (2002). 'The case against intellectual property', *American Economic Review Papers and Proceedings,* 92, 209-212.

Bottazzi, L. and G. Peri (2003). **'**Innovation and spillovers in regions: Evidence from European patent data', *European Economic Review*, 47(4), 687-710.

Branstetter, L. (2001). 'Are knowledge spillovers international or intranational in scope? Microeconometric evidence from the U.S. and Japan', *Journal of International Economics*, 53(1), 53–79.

Chiou, J. Y. (2006), *Essays on the economics of the patent system*, Ph.D. Thesis, University of Toulouse.

Cincera, M. (1997). 'Patents, R&D, and technological spillovers at the firm level: Some evidence from econometric count models for panel data', *Journal of Applied Econometrics*, 12(3), 265-80.

Cohen, W.M., Nelson, R.R. and J.P. Walsh (2000). 'Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or nNot)', W. P. No. 7552, National Bureau of Economic Research.

Crepon, B. and E. Duguet (1997a). 'Estimating the innovation function from the patent numbers: GMM on count panel data', *Journal of Applied Econometrics*, 12(3), 243-263.

Crepon, B. and E. Duguet (1997b). 'Research and development, competition and innovation', *Journal of Econometrics*,79, 355-378.

Cugno, F. and E. Ottoz (2006). 'Trade secret vs. broad patent: The role of licensing', *Review of Law & Economics*, 2(2), Article 3.

Dasgupta, P. and J. Stiglitz (1980) 'Uncertainty, industrial structure, and the speed of R&D', *Bell Journal of Economics*, 11(1), 1-28.

Denicolò, V. (1996). 'Patent races and optimal patent breadth and lenght', *Journal of Industrial Economics*, 44(3), 249-265.

Denicolò, V. (1999). 'The optimal life of a patent when the timing of innovations is stochastic', *International Journal of Industrial Organization*, 17, 827-846.

Denicolò V. and L. A. Franzoni (2004). 'Patents, secrets, and the first inventor defense', *Journal of Economics and Management Strategy*, 13, 517-538.

Duffy, J. (2005). 'A minimum optimal patent term', mimeo, University of California Berkeley.

Duguet, E. and I. Kabla (1998). 'Appropriation strategy and the motivations to use the patent system: An econometric analysis at the firm level in French manufacturing', *Annales d'Economie et de Statistique*, 49/50, 289–327.

Farrell, J. and C. Shapiro (2005). 'How strong are weak patents?', mimeo, University of California Berkeley.

Friebel, G., A.K. Koch, D. Prady and P. Seabright (2006). 'Objectives and incentives at the European Patent Office', mimeo, IDEI, Toulouse.

Gallini, N. (1992). 'Patent policy and costly imitation', *RAND Journal of Economics*, 23(1), 52-63.

Gallini, N. (2002). 'The economics of patents: Lessons from recent U.S. patent reforms', *Journal of Economic Perspectives*, 16(2), 131-154.

Gilbert, R. and C. Shapiro (1990). 'Optimal patent breadth and lenght', *RAND Journal of Economics*, 21(1), 106-112.

Grabowski, H. and J. Vernon (2000). 'Effective patent life in pharmaceuticals', *International Journal of Technology Management*, 19, 98–120.

Griffith, R. (2000). 'How important is business R&D for economic growth, and should the government subsidise it?', IFS Briefing Notes, BN12.

Grossman, G.M. and E.L.C. Lai (2004). 'International protection of intellectual property', *American Economic Review*, 94(5), 1635-1653.

Guellec, D. and Van Pottelsberghe De La Potterie (2003). 'The Impact of public R&D expenditure on business R&D', *Economics of Innovation and New Technology*, 12(3), 225-243.

Guo, J.Q. and P.K. Trivedi (2002). 'Flexible parametric models for long-tailed patent count distributions', *Oxford Bulletin of Economics and Statistics*, 64 (1), 63–82.

Hall, B. H. (1996). 'The private and social returns to research and development: What have we learned?', in *Technology, R&D, and the economy*, edited by B. Smith and C. Barfield, Washington DC, The Brookings Institution.

Hall, B. H., Griliches, Z. and J. A. Hausman (1986). 'Patents and R&D: Is there a lag?' *International Economic Review,* 27(2), 265-283.

Hall, B. H. and R.H. Ziedonis (2001). 'The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry', *RAND Journal of Economics*, 32 (1), 101-128. .

Harhoff, D., F. Narin, F. M. Scherer and K. Vopel (1999). 'Citation frequency and the value of patented innovation', *Review of Economics and Statistics*, 81(3), 511-515.

Hausman, J. A., Hall, B. H. and Z. Griliches (1984). 'Econometric models for count data and with application to the patents-R&D relationship', *Econometrica,* 52(4), 909-938.

Heller, M. and R. Eisenberg (1998). 'Can patents deter innovation? The anticommons in biomedical research', *Science*, 280, 698-701.

Hoti, S., McAleer M. and D. Slottje (2006). 'Intellectual property litigation activity in the USA', *Journal of Economic Surveys*, 20(4), 715-729.

Irmen, A. and M. Hellwig (2001). 'Endogenous technical change in a competitive economy', *Journal of Economic Theory*, 101(1), 1-39.

Jaffe, A. B. and J. Lerner (2004). *Innovation and its discontents: How our broken patent system is endangering innovation and progress, and what to do about it*, Princeton, N.J., Princeton University Press.

Jones, C. I. and J. C. Williams (1998). 'Measuring the social return to R&D', *Quarterly Journal of Economics*, 113, 1119-1135.

Jones, C. I. and J. C. Williams (2000). 'Too much of a good thing? The economics of investment in R&D', *Journal of Economic Growth*, 5(1), 65-85.

Lemley, M. A. (2001). 'Rational ignorance at the patent office', *Northwestern University Law Review*, 95(4), 1495-1529.

Lemley, M. A. and C. Shapiro (2007). 'Patent hold-up and royalty staking', forthcoming in *Texas Law Review*.

Lerner, J. (2002). '150 years of patent protection', *American Economic Review Papers and Proceedings*, 92(2), 221-225.

Lerner, J. and J. Tirole (2004). 'Efficient patent pools', *American Economic Review*, 94(3), 691-711.

Levin, R.C., Klevorick, A.K., Nelson R.R. and S.G. Winter (1987). 'Appropriating the returns from industrial research and development', *Brookings Papers on Economic Activity*, 1987(3), 783-820.

Lichtenberg, F. R. and T. J. Philipson (2002). 'The dual effects of intellectual property regulations: Within- and between-patent competition in the U.S. pharmaceuticals industry', *Journal of Law and Economics*, 45, 643–672.

Loury, G. C. (1979). 'Market structure and innovation', *Quarterly Journal of Economics*, 94(2), 395-410.

Machlup, F. and E. Penrose (1950). 'The patent controversy in the Ninethteen Century', *Journal of Economic History*, 10(1), 1-29.

Mansfield, E., Schwartz M. and S. Wagner (1981). 'Imitation costs and patents: An empirical study', *Economic Journal*, 91, 907-918.

McCalman, P. (2001). 'Reaping what you sow: An empirical analysis of international patent harmonization', *Journal of International Economics*, 55, 161-86.

Montalvo, J. G. (1997). 'GMM estimation of count panel data models with fixed effects and predetermined instruments', *Journal of Business and Economic Statistics*, 15, 82-89.

Nordhaus, W. (1969). *Invention, growth and welfare*, Cambridge, Mass., MIT Press.

O'Donoghue, T. and J. Zweimuller (2004). 'Patents in a model of endogenous growth', *Journal of Economic Growth*, 9(1), 81-123.

Pakes, A. and Z. Griliches (1980). 'Patents and R&D at the firm level: A first look', *Economics Letters*, 5(4), 377-381.

Pakes, A. and Z. Griliches (1984). 'Patents and R&D at the firm level: A first look', in *R and D, patents and productivity*, edited by Z. Griliches. Chicago, University of Chicago Press, 55-72.

Peri, G. (2005). 'Determinants of knowledge flows and their effect on innovation', *Review of Economics and Statistics*, 87 (2), 308-322.

Sain-Paul, G. (2004). 'Are intellectual property rights unfair?' *Labour Economics*, 11(1), 129-144.

Schankerman, M. (1998). 'How valuable is patent protection? Estimates by technology field', *RAND Journal of Economics,* 29(1), 77-107.

Scotchmer, S. (1991) 'Standing on the shoulders of giants: Cumulative research and the patent law', *Journal of Economic Perspectives*, 5(1), 29-41.

Scotchmer, S. (2004), *Innovation and incentives*, Cambridge, Mass., MIT Press.

Shapiro, C. (2001). 'Navigating the patent thicket: Cross licenses, patent pools, and standard-setting', in *Innovation Policy and the Economy*, edited by A. Jaffe, J. Lerner and S. Scott, National Bureau of Economics.

Shapiro, C. (2006). 'Injunctions, hold-up, and patent royalties', mimeo, University of California Berkeley.

Walsh, J.P., W.M. Cohen and A. Arora (2003). 'Patenting and licensing of research tools and biomedical innovation', in *Innovation in a knowledge-based economy*, edited by S. Merrill, R. Levin and M. Meyers, Washington, National Academies Press.

Wang, P., Cockburn, I. M. and M. L. Puterman (1998). 'Analysis of patent data – A mixed-Poisson-regression-model approach', *Journal of Business and Economic Statistics*, 16, 27–41.

SUMMARY. *Is the current level of patent protection too high or too low? To address this issue, this paper reformulates the theoretical analysis of the optimal level of patent protection to take into account the empirical findings of the innovation production function literature. This literature finds a strong relationship between R&D spending and inventions and estimates an elasticity of the supply of inventions of 0.5 or more. The paper the assesses the current level of patent protection, exploiting estimates of the private and social returns to R&D taken from the empirical literature and other available sources. Although more research is needed for a more precise assessment, the evidence available suggests that patents do not over-compensate innovators.*