Do patents over-compensate innovators?

by Vincenzo Denicolò
Università di Bologna[1]

Abstract. I develop a simple and flexible model of the optimal strength of patent protection and use available estimates of the relevant parameters to ask whether patents over-compensate innovators. To this end, I define the strength of patent protection as a profit ratio, namely the ratio of discounted profits actually captured by the patentee to the hypothetical discounted profits he would obtain with infinitely lived, complete monopoly control over the use of his invention. In the baseline model of stand-alone innovations, the profit ratio should be equal to the elasticity of the supply of inventions with respect to R&D expenditures. The empirical literature that has estimated the elasticity of the supply of inventions suggests that a reasonable range for the elasticity is from .5 to .7. In reality, it seems quite unlikely that the representative patentee gets more than two thirds of the hypothetical profits he would get with complete, infinitely lived monopoly control over his invention; probably the representative patentee obtains less than a half of those profits. Moving beyond the baseline model, I discuss a number of additional effects that influence the optimal level of protection. Some of these effects are difficult to assess empirically, and more empirical research is needed to provide a more precise assessment of the over-compensation hypothesis. However, a preponderance of what evidence is available suggests that the representative patentee is not over-compensated.

[1] Mailing address: Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2, I-40126 Bologna, Italy. E-mail vincenzo.denicolo@unibo.it
This paper has been prepared for the April 20-21 2007 Panel Meeting of *Economic Policy* in Frankfurt. I am grateful to Laura Bottazzi, Lorenzo Casaburi, Emmanuel Duguet, Luigi Franzoni, Christine Halmenschlager, Andrea Ichino, Mark Lemley, Steven Martin, Enrico Santarelli, seminar audiences at Kiel, Toulouse, Nancy, Paris and Milan, Editor Paul Seabright and three referees for helpful comments. The usual disclaimer applies.

1. INTRODUCTION

Economists have long recognized that patent design involves a trade-off of fast innovation and static allocative distortion. This trade-off is well understood in theory, but in practice its optimal resolution has proved intractable. Since Nordhaus (1969), the economics literature on optimal patent design has invariably found that the optimal level of patent protection depends on a number of parameters, some of which are very difficult to measure empirically. No wonder that legal scholars came to the pessimistic conclusions that "*this literature has taught us almost nothing* […] *there is little hope that economic analysis can resolve the question of the appropriate scope of the protection of intellectual property*," (Priest, 1986) or "*the literature on the economic effects of patents is especially inconclusive*" (Landes and Posner, 2003).

The goal of this paper is to make some progress toward an analysis with firmer empirical base. I do not report on new empirical work, but I reformulate the theoretical analysis of the optimal level of patent protection so as to highlight the key role of a single parameter, namely, the elasticity of the supply of inventions with respect to R&D expenditures. Starting from Pakes and Griliches (1980), this elasticity has been estimated in several empirical studies, often as a by-product of research targeted at different issues. Since the elasticity of the supply of inventions may vary with the investment in research, the observed elasticity cannot be used to infer the optimal level of patent protection. Nevertheless, it can be used to calculate an appropriate benchmark to test whether patents currently over or under-compensate innovators. [2]

To give up the punch line of the paper, a preponderance of the evidence suggests that patents do not over-compensate innovators. Needless to say, since patent policy is not tailored to individual innovations (one size fits all), certain innovators are inevitably over-compensated – and others under-compensated. In the aggregate, however, the over-reward hypothesis does not seem to be supported by what evidence is available.

A brief preview of how I arrive at this conclusion may help clarify its limits. I define the level or strength of protection as a profit ratio, i.e., the ratio of the discounted profit actually captured by the patentee to the hypothetical discounted profit he would obtain with infinitely lived, complete monopoly control over the use of the new technological knowledge. With this definition, I develop a simple and flexible model of the optimal strength of patent protection that captures the basic

[2] Various papers have compared the private and the social rate of return from innovation, but such a comparison does not really provide a test of the over and under-reward hypotheses. Because patents are distortionary, it is socially costly to use them to incentivize investment in research: thus, the social returns from innovation *should* exceed the private returns at the optimum. Suggestive as it may be, the empirical evidence that the socially optimal investment in research is two to four times actual investment (Jones and Williams, 1998) is therefore consistent with both hypotheses.

trade-off between innovation and monopoly distortions. In the baseline model of stand-alone innovations, *the profit ratio should be equal to the elasticity of the supply of inventions.* Then, I review the empirical literature that has estimated the elasticity of the supply of inventions. Although available estimates vary considerably, this literature suggests that a reasonable range for the elasticity is from .5 to .7. This means that innovators are unlikely to be over-rewarded if they get less than a half, and perhaps even if they get more than a half but less than two thirds of the discounted profits they would get under infinitely lived, complete monopoly over their inventions.

With this benchmark, I then try to assess whether the level of patent protection is currently too high, too low, or is indeed in the optimal range. The measurement of the profit ratio is admittedly a very big problem. The profit ratio depends on a number of technological details, legal rules and policy decisions made by governments, patent offices, and the courts. The policy variable that most obviously affects the profit ratio is the length of protection. The statutory patent life expressed in calendar time is easy to measure; currently, the patent term is 20 years in most countries. However, effective life can differ from statutory life and the choice of the discount rate plays a crucial role in translating calendar-time life into a suitable profit ratio. In addition, "breadth" – that can be loosely defined as the extent to which others are effectively prevented from exploiting patented innovative knowledge for commercial purposes – is arguably more important and certainly more difficult to measure than length. In spite of all these measurement problems, it seems quite unlikely that in reality the representative patentee gets more than two thirds of the discounted profits he would get with complete, infinitely-lived monopoly control over the use of his innovation. Drawing on the limited and sparse available evidence and back-of-the-envelop calculations, my guess is that the profit ratio is probably less than a half; perhaps even a third might be considered a somewhat optimistic estimate. Thus, from the analysis of the baseline model and a review of the empirical literature I take away a feeling that the over-reward hypothesis is not supported by the data.

Moving beyond the baseline model, additional parameters come into play. In developing the baseline model I have tried as far as possible to make assumptions that lead to a conservative calculation of the optimal strength of protection. For example, the baseline model assumes that there are no technological spillovers, that consumers do not benefit from the innovation until after the patent expires, and that competition in research magnifies the incentivizing effect of prospective profits. Relaxing these assumptions raises the benchmark and so reinforces the feeling one takes away from the analysis of the baseline model even in the wake of precise quantitative estimates of the additional parameters. I also argue that with cumulative or complementary innovations the profit ratio should be greater than in the baseline model because of the positive externality that firms racing for an innovation exert on firms racing for complementary or subsequent innovations. Finally, I argue that the fact that innovators can often rely on other mechanisms, such as secrecy or lead time, to appropriate the returns from their investment in research does not mean that patent protection should be weakened, as long as those mechanisms are also distortionary.

However, some of the assumptions of the baseline model in fact inflate the benchmark strength of protection. In particular, the baseline model abstracts from transaction costs, business stealing, and opportunistic behaviour on the part of firms seeking patent protection. I discuss each of these effects, showing how it impacts the benchmark strength of protection. Unfortunately, while the importance of these effects is undeniable, their size is largely unknown. Therefore, any policy conclusion is inevitably tentative at this stage, and more empirical evidence is needed to provide a more precise assessment of the over-reward hypothesis.

The rest of the paper is organized as follows. The next section situates the paper in a fuller account of current policy concerns. Section 3 develops the baseline model. Section 4 reviews the empirical literature and tries to assess the current level of patent protection. Section 5 relaxes the assumptions of the baseline model that deflate, or have an ambiguous impact on, the benchmark profit ratio. Section 6 discusses the assumptions that inflate the benchmark. Section 7 summarizes the main results of the paper and offers some concluding remarks.

2. IMPLICATIONS FOR PATENT POLICY

What policy implications can and what cannot be drawn from an assessment of the under- and over-compensation hypotheses? At the most basic level, such an assessment would provide a preliminary test of the recurrent proposal of abolishing intellectual property straightway – a proposal that is based on the explicit or implicit claim that innovators are substantially over-rewarded. More generally, an assessment of whether patents under- or over-compensate innovators would offer invaluable guidance on policy by indicating the proper direction for patent reform. However, not even a precise assessment would automatically solve many of the currently debated patent policy issues.

To begin with, it is sometimes difficult to tell whether a specific policy move raises or lowers the profit ratio. Moreover, patent design involves problems that go beyond the size of the pie to be granted to innovators. First, when innovation is sequential, or several innovative components must be assembled together to operate a new technology, patent policy must determine how the pie should be divided among all firms that concurred to the discovery. If certain innovators get more and others less than their appropriate share, incentives to innovate can be seriously distorted. Second, for any given size of the pie, policymakers often have some latitude on how to prepare it. The policy tools or combination of tools used to reward innovators affect the deadweight loss society bears per unit of profit it provides to innovators, and thus ultimately determine the terms of the trade-off between innovation and monopoly distortions. Improving the terms of the trade-off may actually be more important than resolving the trade-off optimally.

Those caveats must be borne in mind when evaluating actual or proposed policy moves. Consider for example the upward trend in patent protection that has characterized the last quarter of a century. In most developing countries and also in some industrialized countries, the adoption of the TRIPs agreement required major reforms that extended the coverage of patent protection to the pharmaceutical, food and chemical industries, strengthened the enforcement of patents, and extended the patent term to 20 years (McCalman, 2001). In the US, the creation in 1982 of new Courts of Appeals for the Federal Circuit (CAFC) providing a unified judicial appellate authority for all patent cases resulted in broader interpretation of the "doctrine of equivalents" [3] and a rise in damages awarded for patent infringement (Jaffe, 2000). These changes probably resulted in an increase in the profit ratio, and so they would be broadly desirable if currently innovators were not over-compensated. However, other tools could have been used to achieve the same goal. For instance, one might argue that extending the patent term in developed countries is preferable to expanding patent protection in developing countries, or lowering litigation costs is better than raising damages awarded for patent infringement.

Other policy changes that are often described as moves towards stronger patent protection may actually not raise the profit ratio. For example, the American patent system has experienced an unprecedented broadening in the realm of patentability. There are now thousands of patents for living organisms, genetic sequences, software programs, business methods etc., while none of these inventions would have been patentable in the 1970s. Have these patents resulted in a higher profit ratio? The answer is highly uncertain. Consider for example gene patents. Often, decoding genetic sequences is instrumental to the search for new drugs. However, since new drugs have long been patentable, making genetic sequences also patentable does not necessarily raise the profits from the sale of new drugs. To the contrary, making genetic sequences patentable may create a problem of Cournot complements[4] or raise transaction costs thereby reducing the profit ratio in the pharmaceutical industry. In addition, and perhaps more importantly, making genetic sequences patentable may alter the division of profit between upstream and downstream firms in the pharmaceutical industry. This may change the relative incentives to conduct upstream and downstream research, with a final effect on social welfare that is quite uncertain.

---

[3] The doctrine of equivalents is a legal rule that allows a court to find a patent infringement even though the infringing device or product does not fall within the literal scope of the patent's claims.

[4] 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 thereby reducing aggregate profits below monopoly profits (Shapiro, 2001).

Similar remarks apply, for instance, to software patents. More patents, that is to say, are not synonymous with stronger patent protection. This is all the more true when the increase in the number of patents results from a weakening in the novelty and non-obviousness standards for patentability. Granting patents on technological knowledge that is already in the public domain or is obvious may give opportunistic agents the power to tax not only consumers but also genuine inventors, whose incentives to innovate are correspondingly reduced. In the US, the Patent Office and the courts have apparently gone so far in weakening the novelty and non-obviousness requirement that many practitioners are concerned that the patent system is broken and in need of repair.[5] Needles to say, a finding that patents do not over-compensate innovators would not dispel these concerns.

To summarize, patent policy is multi-facets and this (already too long) paper is silent on many important issues. A finding that patents do not over-compensate innovators would not dispel concerns that the realm of patentability may have been excessively broadened, that the patentability requirements may have been unduly weakened, that certain innovators may get more than their fair share of the profits arising from a complex technology that involves several innovative components, or that we may not be using the most efficient tools to encourage innovation.

Having determined the scope of the analysis, we can set at work.

3. BASELINE MODEL

The goal of this section is to develop a parsimonious and as far as possible conservative characterization of the optimal level of patent protection. This requires a careful choice of modelling assumptions.

Throughout the paper, I take the nature and the size of the innovation as fixed. I denote by π the (flow) profit that would be earned by a fully protected innovator, by D the corresponding deadweight loss, and by $V := \pi + D + CS$ the (flow) social value of the innovation when the innovation is in the public domain. Here CS is a residual. For example, in the case of product innovation described in Box 1, CS is consumer surplus. More generally, however, CS may capture technological spillovers and other positive externalities. To keep the analysis as general as possible, I treat π, D, and CS as exogenous parameters, without making any specific assumption on the nature of the innovation, demand, and the competition in the product market.

---

[5] Europe seems relatively immune from some of the alleged defects of the American patent system. For example, gene patents are much tougher to uphold in Europe than in the US, and business methods and software patents are still controversial. Moreover, the European Patent Office seems to grant less dubious patents than the USPTO, although it also suffers from problems of overload and lack of appropriate incentives (Friebel *et al*, 2006).

Assuming a stationary environment, the social value of the innovation is then $v^S := \frac{V}{r}$, where r is the (private and social) discount rate.

The analysis that follows focuses on two policy variables, the length and 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. For any given r, it ranges from 0 ($T = 0$) to 1 ($T = \infty$). To capture the fact that patents are limited in breadth, following Gilbert and Shapiro (1990) I denote by β the fraction of the flow profits π that the innovator earns while the patent is in force; the remaining fraction $(1-\beta)$ spills over to somebody else (e.g. consumers or other firms). When breadth is limited, I assume that the deadweight loss is proportionally reduced, i.e. βD. Later I shall discuss the consequences of relaxing this assumption.

Box 1: Drastic and non-drastic innovations

For illustrative purposes, Figure 1 depicts a textbook example of a product innovation. The good cannot be produced before the innovation. Once the innovation is achieved, the patentee charges the monopoly price p_M until the patent expires. In this phase the flow social benefit from the innovation is $\pi + CS$. When the patent expires, the price falls to marginal cost c_1 and society nets a (flow) benefit $V = \pi + D + CS$.

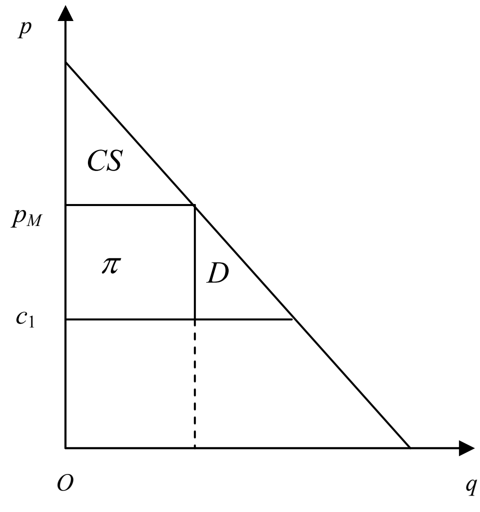

Fig. 1: product innovation.

The case of a drastic, cost-reducing or quality-improving innovation is similar. By way of contrast, Figure 2 depicts the case of a non-drastic innovation.

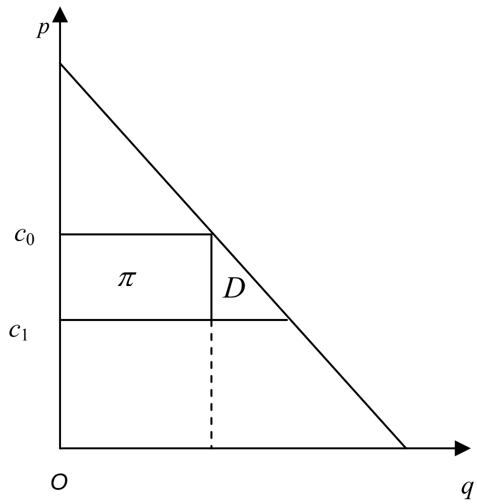

Fig. 2: non-drastic, cost-reducing innovation.

Here the innovation reduces the unit production cost from the pre-innovation level c_0 to $c_1 < c_0$, where 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 social flow value of the innovation. Here $CS = 0$ (note, CS denotes the *increase* in consumer surplus) and therefore $V = \pi + D$. This is the case I focus on in the baseline model. In this case, customers do not benefit from the innovation until after the patent expires.

Summarizing, with infinitely lived protection ($z = 1$) and maximum breadth ($\beta = 1$) the innovator's discounted profits would be $\frac{\pi}{r}$. When protection is limited in scope and over time, his profits are only $v^P := \beta z \frac{\pi}{r}$. Therefore, normalized length z and breadth β combine into an overall profit ratio, βz, which indexes the level or strength of protection. In words, βz is the ratio of profits actually captured by the patentee to the hypothetical profits that he would obtain with infinitely lived, complete monopoly control over the use of the new knowledge.

Box 2: Assumptions of the baseline model

The following assumptions of the baseline model will be relaxed later:
- a stand-alone innovation with no subsequent developments or improvements;
- the innovation does not require complementary knowledge that is in turn protected by separate intellectual property rights;
- innovative knowledge is completely non-excludable in the absence of patent protection;
- the innovator's profits are driven to zero as soon as the patent expires;
- there are no technological spillovers and consumers do not benefit from the innovation until after the patent expires;
- free entry in research;
- the deadweight loss increases linearly with patent breadth;
- zero transaction costs;
- no business stealing.

To proceed, I use a simple reduced-form model of innovative activity. The innovation can be achieved instantaneously with an aggregate probability x at a cost $\alpha c(x)$. The aggregate R&D expenditure function $c(x)$ is increasing and convex: $c'(x) > 0$ and $c''(x) > 0$, $c(0) = 0$, and α is a positive parameter that captures the difficulty of achieving the innovation. To illustrate, assume that research firms can operate a number of indivisible, uncorrelated research projects. 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 succeeds, namely $x = 1 - (1 - p)^n$, and the aggregate R&D cost is nq. Since $n = \frac{\log(1-x)}{\log(1-p)}$, the aggregate R&D cost can be rewritten as $\alpha c(x)$ where $\alpha = -\frac{q}{\log(1-p)} > 0$ and $c(x) = -\log(1-x)$. It can be easily checked that this function is increasing and convex and satisfies the condition $c(0) = 0$.

Finally, in the baseline model I assume that $CS = 0$ and that there is free entry in research which will drive the expected net profits of innovators to zero. As is well known, both assumptions have an unambiguous negative effect on the optimal strength of protection (see Denicolò, 1999; Duffy, 2005), and both will be relaxed below. Box 2 summarizes the assumptions of the baseline model that will be relaxed in subsequent sections.

With these assumptions in place, 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 of research firms, the expected profit of firms investing in research must vanish. The expected profit equals the probability of success, x, times the private value of the innovation, $v^P = \beta z \frac{\pi}{r}$, minus the research costs. To illustrate, continuing with the above example suppose that if two or more projects succeed simultaneously, a project is randomly selected and the patent is granted to the firm running that project. Then, the expected profit from running a project is $\frac{x}{n} v^P - q$, whence it is clear that the zero-profit condition reduces to (1). Technical Appendix A shows that equation (1) can be regarded as a reduced form of many other seemingly different models of investment in research used in the economics literature, including the Poisson model.

The second equation of the baseline model is the social welfare function to be maximized. I use the standard definition of social welfare in a partial equilibrium framework,[6] i.e. the discounted aggregate benefit from the innovation, net of R&D costs:

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

The meaning of this equation is as follows. The innovation is achieved with probability x. Conditional on success, the flow social value of the innovation is $[\pi + (1-\beta)D]$ before the patent expires, i.e. for a normalized time length z, and $(\pi + D)$ thereafter, i.e. for the remaining time $1 - z$. Note that while the patent is in force, the share β of D is lost due to supra-competitive pricing. 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. Equation (2) neatly illustrates the trade-off of innovation and monopoly distortions. A rise in βz raises the probability of success x *via* an increase in the incentives to innovate, but it also lowers the fraction of v^S that society nets conditional on success, i.e. the "welfare ratio" $(1 - \beta z)$. This

[6] Although it is possible to cast the analysis in a general equilibrium framework, the general equilibrium effects do not substantially alter the terms of the trade-off of innovation and monopoly distortions: compare e.g. Grossman and Lai (2004). In particular, Saint-Paul (2004) asks whether distributional concerns would have the effect of reducing the optimal strength of intellectual property protection, but finds that any such effect is small.

fraction is the complement to one of the profit ratio because a fraction βz of π is dissipated in the patent race and a fraction βz of D is destroyed as a deadweight loss. This implies that in the baseline model the optimal strength of protection is necessarily strictly lower than one.

The policy-maker must choose the length and breadth of protection, z and β, so as to maximize social welfare taking into account that x is implicitly given by the zero-profit condition (1). The solution to the social problem leads to the following simple "elasticity rule" (see Technical Appendix B for a proof):

(3) $$\beta z = \eta$$

where $\eta := \frac{c(x)}{xc'(x)}$. To interpret equation (3), note that η is the elasticity of the probability of success x with respected to R&D investment. With a large number of potential inventions, however, the supply of inventions is proportional to the probability of success x and so η is also the elasticity of the supply of inventions. Then, the elasticity rule simply says that the profit ratio should be equal to the elasticity of the supply of inventions. This result is especially noteworthy because it is extremely simple and highlights the key role of a parameter, the elasticity of the supply of inventions, that has been estimated in several empirical studies.[7] The elasticity rule thus opens the way to an analysis of the optimal level of patent protection with firmer empirical base.

Some intuition for the elasticity rule can be gained as follows. A one percent increase in the profit ratio βz raises the innovator's expected profit by one percent. By the zero-profit condition, this raises R&D expenditures by one percent, which in turn raises the supply of inventions by η of one percent. This, however, is just the direct effect of a rise in the profit ratio on the supply of inventions. If the probability of success raises by η of one percent, the expected profit, and hence R&D expenditures, will in turn increase by the same amount. This second-round effect raises the number of inventions by η times η of one percent, i.e. η^2 of one percent. But this further raises the expected profit, with a multiplicative effect that is similar to a Keynesian multiplier. The total effect of a one percent increase in the profit ratio on the supply of inventions is therefore $\eta/(1-\eta)$ of one percent. Now, at the optimum the

---

[7] Although several papers have analyzed the trade-off of innovation and monopoly distortions at the theoretical level, so far the role of the elasticity of the supply of inventions has been rather overlooked. Some papers implicitly assume specific numerical values for 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 other parameters. See e.g. Duffy (2005) for a recent contribution that shares some of the assumptions of the baseline model.

percentage increase in the supply of inventions associated with a one percent increase in the profit ratio must equal the corresponding percentage decrease in the welfare ratio $(1-\beta z)$. Since the elasticity of the welfare ratio $(1-\beta z)$ with respect to the profit ratio βz is simply $\beta z / (1-\beta z)$, the elasticity rule follows by simply equating $\eta / (1-\eta)$ and $\beta z / (1-\beta z)$.

Several further comments are in order. First, the elasticity η is necessarily lower than one given the assumption that the function $c(x)$ is convex and $c(0) = 0$. This assumption captures the notion that the creation of new knowledge requires at least one input in fixed supply. Such a fixed input may be talent, it may be the set of good ideas at any given point in time, or it may be something we do not fully understand yet, but it is difficult to imagine that doubling the amount of resources invested in research doubles the output – if the output is measured properly.

Second, for any given profit ratio βz, the combination of length and breadth is a matter of indifference, both for the innovator and for society. This result is due to the assumption that the deadweight loss from patent protection is proportional to the breadth of protection β. If this proportionality assumption fails, a non trivial problem of choosing the optimal combination of patent length and breadth arises (see section 5.3 below).

Third, various parameters of the model that *prima facie* should affect the optimal strength of protection, like π, D and α, do not enter the formula for the profit ratio directly. In particular, the optimal strength does not directly depend on the size of the innovation or on the size of the market, as measured by $\pi + D$. The intuitive reason is the following. A rise in the social value of the innovation $\pi + D$ raises the investment in research that society ought to make to achieve that innovation. However, for any given level of the profit ratio, a rise in π raises the private value of the innovation and so it automatically raises the equilibrium level of investment in research. In the baseline model, there is no need of adjusting the strength of protection. For similar reasons, the optimal strength of protection does not directly depend on the difficulty of achieving the innovation as captured by the shift parameter α. A rise in α lowers both the socially optimal and the equilibrium level of investment in research, with no need of adjusting the strength of protection.

However, a change in the value of the innovation or the difficulty of obtaining it can change the optimal strength of protection to the extent that it affects the elasticity η. If η increases (resp., decreases) with the level of investment in research, a rise in the size of the market or in the size of the innovation raises (resp., lowers) the optimal strength of protection by raising the equilibrium level of R&D expenditure. The same is true of a fall in the difficulty of achieving the innovation. In general, it is hard to tell whether the elasticity of the supply of inventions raises, falls or stays constant with the level of investment in research,[8] an issue which is however largely irrelevant to our purposes.

The fact that the optimal strength of protection does not directly depend on the value of the innovation allows an immediate extension of the elasticity rule (3) to the case of heterogeneous innovations. With a distribution of innovations that differ in value or R&D cost, equation (3) continues to apply with no change 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, with weights that reflect the values of different innovations. Moreover, heterogeneity raises the benchmark strength of protection in that a mean preserving spread in the distribution of the individual elasticities raises the optimal value of βz (see Technical Appendix B for details).

Finally, the elasticity rule asserts that the optimal strength of protection does not depend on the ratio D/π. This result might seem surprising since the ratio D/π measures how costly it is to incentivize investment in research and therefore determines the terms of the trade-off between innovation and monopoly distortions. The intuition for this result is clearest if we focus on the case of maximum breadth (a similar argument however holds when $\beta < 1$). With $\beta = 1$, any net social benefits from the innovation start accruing only after the patent expires. Before the patent expires, 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. After the patent expires, there is no longer any deadweight loss and society enjoys the entire social value of the innovation, $V = \pi + D$. Therefore, the ratio D/π does not affect the optimal level of protection. Of course, this does not mean that it is not important to minimize static allocative inefficiencies while the patent is in force. For any given V and any given fixed profit ratio, any policy move that lowers D and raises π raises the market equilibrium investment in research thereby increasing social welfare.

The fact that the optimal strength of protection is lower than one even when $D = 0$ may seem particularly surprising. When $D = 0$, in the baseline model there are no social costs associated with patent protection, and therefore it would seem that setting $\beta z = 1$ is desirable. However, recall that with competition in research the winner-take-all effect leads to over-investment. Limiting the strength of protection ameliorates this over-investment problem. On the other hand, I shall show below that with monopoly in research and $D = 0$ it is indeed optimal to set $\beta z = 1$. This illustrates the more general principle that the optimal strength of protection decreases with the intensity of competition in research.

---

[8] Boldrin and Levine (2005) implicitly argue that the elasticity of the supply of inventions necessarily falls for extremely large values of the investment in research. Although theirs is an asymptotical result, it may be taken to mean that the elasticity is decreasing over the relevant range.

4. PRELIMINARY ASSESSMENT

In this section I turn from theory to empirical evidence. I review empirical estimates of the elasticity of the supply of inventions with respect to R&D expenditures. This parameter, that fully determines the optimal strength of protection in the baseline model and plays a major role even in more highly structured models, has indeed been estimated in several empirical studies.[9] More tentatively, I also try to assess the strength of patent protection in current practice. This will lead to a preliminary assessment of whether patent protection is excessively strong.

4.1. Empirical estimates of the elasticity of the supply of inventions

What is the ideal evidence to estimate the elasticity of the supply of inventions? In an ideal experiment, one would observe two groups of firms, one investing in R&D more than the other for some exogenous reason, that are otherwise identical. One could then calculate the elasticity of the supply of inventions by looking at how many more inventions firms in the high-investment group achieve as compared to firms in the low-investment group.

Of course, this ideal experiment is never observed in practice. We do have data on different firms and regions in the world, but these firms and regions differ in many dimensions, R&D expenditures do not vary exogenously, and the number of inventions must often be proxied by the number of patents or patent applications. Using this data, econometricians can at best approximate the ideal experiment.

The most creative approximation to date is perhaps found in Linn and Acemoglu (2004). They focus on the pharmaceutical industry, where the number of newly approved drugs (as opposed to the number of patented new chemical substances, many of which fail to pass the pre-clinical and clinical tests) constitutes a fairly precise measure of the number of innovations. They do not look at R&D expenditures, but directly measure the impact of changes in the size of the market, that in their sample varies exogenously for demographic reasons, on the output of innovative activity. Proceeding this way, 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 findings imply an elasticity of the supply of inventions ranging from .8 to .85.[10]

[9] Although the theoretical analysis developed in this paper applies also to copyrights, I focus on patents because to the best of my knowledge there are no empirical analyses of the elasticity of the supply of copyrightable works.

[10] In this paper's notation, the elasticity empirically estimated by Linn and Acemoglu is $\frac{dx}{dv^P}\frac{v^P}{x}$. One can easily show that $\frac{dx}{dv^P}\frac{v^P}{x} = \frac{\eta}{1-\eta}$.

Turning to more standard evidence, several empirical studies have tried to estimate an "innovation production function"

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

that relates the number of patent applications at time t, P_t, to current and past R&D expenditures, $R_t, R_{t-1}, ...$ Apart from the time lags and the use of *P* as a proxy for *x*, the function $f(.)$ is just the inverse of our R&D expenditure function $\alpha c(x)$. Therefore, with a Cobb-Douglas specification of the innovation production function, the estimated parameters can be directly interpreted as the appropriate elasticities.

Using a 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 .61. Their analysis has been subsequently extended and refined by Hausman *et al* (1984) and Hall *et al* (1986). They note that a log-linear regression model is inappropriate for firm-level data since many observations have zero patents, and suggest 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 .87 using the Poisson distribution and an elasticity of .75 using the negative binomial distribution. However, their data suggests the existence of strong firm-specific effects. Accounting for firm-specific effects lowers the estimated elasticities: for example, with the negative binomial model the total elasticity of .75 falls to .52 with random effects and to .37 with fixed effects. However, Hausman *et al* (1984) note that firm-specific effects may actually reflect pre-sample R&D, and conclude that while the elasticity of the supply of inventions to current R&D is less than .4, the total elasticity (including lagged and pre-sample R&D) is "at least .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 expenditures and patenting, with an elasticity of about .3; that the contribution of observed past R&D is on the order of .05; and the contribution of unobserved or pre-sample R&D is about .25 (p. 281). These effects add up to a total elasticity of the supply of inventions of about .6.

A similar pattern emerges also 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 of the supply of inventions are typically large, around .8-1, but the estimations are much lower once individual heterogeneity is somehow taken into account. For example, using quasi-differenced

GMM estimation, Cincera (1997) finds an elasticity of .48 and Crépon and Duguet (1997a) an estimate as low as .26.

Blundell *et al* (2002) suggest a possible resolution to this apparent inconsistency. They note that the process under study here 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 account of unobserved heterogeneity are upward biased, but estimators commonly used to account for firm-specific effects are seriously downward biased. They also propose an alternative "pre-sample mean" estimator that replaces the fixed effect by the pre-sample mean of the dependent variable – here, the number of patents. This estimator performs well in the Monte Carlo simulation study, and when applied to the NBER data set delivers a point estimate of the elasticity η of about .51 (Blundell *et al*, 2002).

More recent work exploits survey data in addition to patent data to explicitly model the decision to patent. 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 is around .6-.7, but in the structural form model it falls to .4-.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 endogenously determined. In estimating the elasticity of the supply of inventions, they also condition on 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 .61, both in the single equation and in the system estimates.

Another strand of the literature (Bottazzi and Peri, 2003, and Peri, 2005) estimates the elasticity of the supply of inventions using regional-level rather than firm-level data. By relying on a very large number of patents in each region (e.g., almost 10,000 per region on average in Peri, 2005), these studies at least partially overcome the problem of the variability in the value of innovations since differences in the value of individual patents are likely to be averaged out in large aggregates. Bottazzi and Peri's estimations of the elasticity of the supply of inventions vary from .67 to .96, while Peri's (2005) estimations vary from .6 to .81.

Although no individual piece of evidence is especially compelling, the collection taken together suggests that the elasticity of the supply of inventions may be prudently taken to range between .5 and .7. However, the fact that available estimates vary considerably also suggests that the true elasticity of the supply of inventions may itself vary across sectors and over time. So, a lot of caution should be used in drawing policy conclusions.

In particular, it is important to stress that the elasticity may change with the level of investment in research. Although most empirical studies use a Cobb-Douglas specification of the innovation production function, the few studies that allow for more flexible functional forms find strong evidence that the elasticity is not constant (Guo and Trivedi, 2002). This means that the observed elasticity cannot be used to calculate the optimal strength of patent protection. However, since the elasticity rule (3) holds at the social optimum even if the elasticity is not constant, the observed elasticity provides a proper benchmark to test the over-compensation hypotheses. In the light of the elasticity rule, the empirical literature surveyed here suggests that patentees are not over-rewarded if they get less than a half, and perhaps even if they get more than a half but less than two thirds of the discounted profits they would get under infinitely lived, complete monopoly over their inventions.

4.2. Normalized length

Where do we stand? There seems to be no empirical study of the profit ratio. To make some progress, I address the two components of the profit ratio, the normalized length z and breadth β, separately.

At first sight, length is easy to measure: in most countries, the statutory patent life is 20 years from the date of the first filing of the patent application. However, the effective patent term is frequently less than 20 years because patents are often obtained long before products are actually marketed. This problem is particularly acute in the pharmaceutical sector, where effective patent life has been estimated at less than 12 years in spite of various provisions that extend statutory patent life to compensate for the time lost while developing the product and awaiting regulatory approval (Grabowski and Vernon, 2000).

Next, in order to translate effective patent life measured in calendar time into a profit ratio, an appropriate discount rate must be used. To begin with, the long-run real interest rate should be lowered by the economy's rate of growth. The difference between the long-run, risk-free real interest rate and the economy's rate of growth is probably small, say 1-2%. However, the resulting discount rate should be augmented by the instantaneous probability that the innovation is superseded by subsequent exogenous technical progress that makes the innovation valueless. Only exogenous technical progress should be taken into account here. The reason is that if original patentee (say, the inventor of the antiretroviral therapy against HIV) was granted full monopoly control over his innovation, including the power to block the use of subsequent improvements (Scotchmer, 2004), his profit would not be terminated by the occurrence of subsequent improvements that build on the original innovation (like e.g. me-too or second-generation antiretroviral drugs).[11] However, his profit would be

---

[11] Of course, even me-too or second-generation drugs erode the profits of the original patentee, but this is due to the limits in the scope of patent protection, an effect that in our framework is captured by the fact that $\beta < 1$. See the discussion of patent breadth in the pharmaceutical sector developed below.

terminated by an independent technological breakthrough (like an effective vaccine against HIV). Now, if the arrival of major exogenous technological breakthroughs follows a Poisson process, an instantaneous probability of 4% means that a major breakthrough occurs every 25 years. It seems unlikely that major exogenous breakthroughs be much more frequent. Finally, it is unclear whether a risk premium should also be included. A careful treatment of risk would perhaps require distinguishing between the social and the private discount rate.

Table 1: Normalized patent length *z*

yearly discount rate r / *effective patent length T (in years)*	2%	3%	5%	7%	10%
20	**.33**	**.45**	**.63**	**.75**	**.86**
16	**.27**	**.38**	**.55**	**.67**	**.80**
12	**.21**	**.30**	**.45**	**.57**	**.70**

Table 1 illustrates how normalized length depends on effective patent length and the discount rate. A discount rate of 10% should perhaps be considered unreasonably large. For large but more reasonable values of the discount rate like 5-7%, normalized length lies in the interval 45-75%. This means that the mere fact that patent protection is limited in time reduces the profit ratio considerably below one. However, breadth is arguably even more important than length in limiting the strength of protection.

4.3. Breadth

Patent breadth is limited for three broad sets of reasons. First, the enforcement of patents is largely incomplete. Detecting infringements is difficult, especially when there are many potential users of the new knowledge. Litigation is costly and may be risky, since patents are frequently invalidated by the courts. Since the private and social costs of a perfect enforcement would be enormous, some degree of infringement is inevitably tolerated. Second, patented innovative technological knowledge can be lawfully imitated to some extent even before the patent expires. “Inventing around” a patent is a strategy encouraged by the law and routinely used by competitors to reduce the patentees’ competitive advantage. Even if policymakers intended to draft the patent claims so as to prevent any imitation, the inherent ambiguity of human language would create a risk that the inventor might obtain a monopoly over something that is already known, and as such should remain in the public domain, or over something that has not been invented yet, which would stifle subsequent innovation. In practice it is impossible to avoid those risks without limiting the breadth of protection. Third, patent protection is limited geographically.

In spite of the TRIPs agreement, in many countries patent protection is practically absent.

Although it is clear that in practice β is substantially lower than one, it is difficult to say by how much. The empirical evidence is very sparse and indirect. However, two pieces of evidence suggest that β may be tentatively set at around .5 or even less. First, in a classic study Mansfield *et al* (1981) find that more than 60% of all innovations in their sample, of which more than three quarters were patented, were imitated within 4 years from discovery. Unfortunately this evidence does not suffice to provide an estimate of patent breadth in the absence of further assumptions. Suppose, however, that the stochastic event of successful imitation follows a Poisson process with parameter μ. Here μ is the conditional probability that an innovation is imitated in the current year, given no imitation to date. Mansfield *et al*'s data imply that $\mu = .23$. Next, suppose that "imitation" halves the innovator's profits, since the patentee must now share the market with the imitator. Assume a patent life of 20 years and a discount rate of 5% (the calculation is not very sensitive to the choice of the discount rate, however). Then, the profit ratio is about .32, which is roughly a half of the profit ratio of .63 that corresponds to $T = 20$ and $\beta = 1$ (see Table 1). This implies a value of β of about .5 and is probably a conservative guess. The assumption that imitation only halves the patentee's profit is indeed optimistic, since generally competition from imitators lowers industry profits and there can be multiple imitations over time. More important, the process of inventing around a patent, which Mansfield *et al* focus on, is only one reason why breadth can be limited.

More recently, Lichtenberg and Philipson (2002) have studied the patentees' profit erosion process in the pharmaceutical sector. They distinguish between entry by manufacturers of generic drugs when the patent expires, which they call "within-patent entry," and entry by producers of new competing drugs, which they call "between-patent entry." Thus, within-patent entry reflects the fact that patent protection is limited in length, whereas between-patent entry captures both the fact that patent protection is limited in breadth ("imitative" between-patent entry) as well as independent technical progress ("innovative" between-patent entry). Unfortunately, Lichtenberg and Philipson do not provide separate estimates of the effects of "imitative" and "innovative" between-patent entry, although the fact that the majority of newly approved drugs are me-too drugs suggests that "imitative" between-patent entry may be more important than "innovative" between-patent entry. At any rate, Lichtenberg and Philipson find that in the pharmaceutical sector "between-patent entry has about four times as large an effect on [discounted sales] as within-patent entry." Thus, if just one fourth of between-patent competition is imitative, Lichtenberg and Philipson's finding would imply that breadth is as important as length in constraining the patentee's ability to appropriate the returns from the innovation. With an effective patent life of 12 years and a discount rate around 5-7%, in the pharmaceutical sector normalized length z would be around .5. Lichtenberg and Philipson's results would then imply that it is unlikely that β exceeds

.5 in the pharmaceutical sector.[12] Note that this is a sector where patent protection is notoriously considered to be more important than in many other industries in spite of a relatively short effective patent length (Cohen *et al*, 2000), suggesting that here the breadth of protection is particularly large.

4.4. Summary

Available empirical estimates of the elasticity of the supply of inventions suggest that the benchmark profit ratio may range from .5 to .7. In practice, normalized patent length is probably below .75 and the breadth of patent protection is unlikely to exceed .5. Therefore, it seems that the profit ratio for the representative patentee is less than a half, and perhaps even a third is a rather optimistic guess. Thus, from the analysis of the baseline model and a review of the empirical literature I take away a feeling that the over-compensation hypothesis is not supported by the data.

I recognize that this is a very preliminary and tentative assessment with which reasonable people may well disagree. If the elasticity of the supply of inventions was .4, the effective patent life coincided with statutory life, the appropriate discount rate was 7%, and patent breadth was around two thirds, the elasticity rule would imply that patentees are moderately over-compensated. None of these guesses is unreasonable. However, all seem somewhat biased and all are needed to support the over-compensation hypothesis.

5. RELAXING THE ASSUMPTIONS

In this section I relax the simplifying assumptions of the baseline model that are likely to deflate the benchmark strength of protection. The analysis leads to a series of modified elasticity rules, the formal derivation of which is relegated to Technical Appendix C.

5.1. Drastic innovations and spillovers

The baseline model assumes that society does not enjoy any positive net benefit from the arrival of the innovation until after the patent expires. Formally, in the baseline model I have set $CS = 0$. This assumption rules out technological spillovers as well as the possibility that customers benefit from the innovation even before the patent expires (see Box 1). More generally, the parameter CS can capture various sorts of positive externalities associated with innovative activity.

When $CS > 0$, the elasticity rule (3) changes as follows:

(5) $$\beta z = (1+\sigma)\eta$$

[12] Moreover, Lichtenberg and Philipson focus on the US and so they do not account for counterfeiting, a phenomenon that has been estimated to account for nearly 10% of total drug sales at the world level.

where $\sigma := \dfrac{CS}{\pi + D}$. Thus, when $CS > 0$ the benchmark strength of protection is greater than with $CS = 0$ and, more generally, it increases with CS. The intuitive reason is that any positive externality arising from innovative activity calls for stronger incentives to invest in research. The empirical literature that has tried to measure technological spillovers and the gap between the social and private returns from innovation suggests that σ can be large. (In principle, if σ is sufficiently large, the social problem may have a corner solution; in other words, it might become socially desirable to set $\beta z = 1$, since society now gets a positive net benefit from the innovation even before the patent expires.)

5.2. 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 research. This is probably an optimistic assumption. However, it turns out that this assumption leads to a conservative calculation of the optimal strength of patent protection.

To see why, consider the extreme case of monopoly in research. The innovator's expected profits are no longer driven to zero; rather, the monopolist chooses its research effort x so as to maximize its expected profits $xv^P - \alpha c(x)$. With an iso-elastic R&D cost function, the elasticity rule changes into

(6) $$\beta z = \frac{\pi + D}{\eta\pi + D}\eta$$

Since $\eta < 1$, the optimal strength of protection under monopoly is larger than under free entry in research. (In particular, with $D = 0$ the social problem has now a corner solution $\beta z = 1$.)

Although the case of monopoly in research is rather extreme, it illustrates the more general principle that the weaker is the intensity of competition in the invention industry, the greater the benchmark (Denicolò, 1999). The reason is that more competition in research raises the equilibrium R&D investment for any given reward to the innovator v^P, and hence for any given level of the profit ratio. Since the government grants patents in order to encourage innovative activity and strengthening patent protection is socially costly, with more competition in research there is less need of rewarding innovators generously.

5.3. The optimal combination of breadth and length

In the baseline model I have assumed that the deadweight loss is proportional to the breadth of patent protection. While it seems natural to posit that the deadweight loss increases with the breadth of protection, the proportionality assumption is restrictive.

For example, in the model of Gilbert and Shapiro (1990) the deadweight loss is a convex function of patent breadth, while in Gallini (1992) it is concave.

To cover these cases, assume that the deadweight loss borne 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 a matter of indifference. When the function $D(\beta)$ is not affine, a non trivial problem of choosing the optimal combination of patent breadth and length arises. Typically, this problem has a corner solution with either $\beta = 1$ or $z = 1$.[13] Which corner is optimal depends on the sign of the second derivative $D''(\beta)$ (Denicolò, 1996). 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 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 Technical Appendix C shows that the optimal profit ratio is greater than in the baseline model.

5.4. Complementary innovations

In many industries, such as telecommunications, biotechnology, and software, production often requires the use of many complementary innovative technological components. If a single firm controls all of the separate property rights, nothing substantial changes. Often, however, complementary pieces of innovative knowledge are owned by different firms. This generates new problems. First, a proliferation of patents held by different owners might prevent manufacturers from obtaining the right to develop the new products, creating the so-called tragedy of the anti-commons (Heller and Eisenberg, 1998). Second, with complementary intellectual property rights there arises a problem of Cournot complements (Shapiro, 2001). Although firms will naturally try to remedy these problems, static deadweight losses are perhaps larger than in the baseline model of stand-alone innovations. In the extreme case in which the tragedy of the anti-commons completely blocks the development of the new product until after the patents expire, it is as if $\pi = CS = 0$ and $D = V$.

One might think that the larger static allocative inefficiencies associated with complementary patents should lower the optimal level of protection. As it turns out, this intuition is misleading.[14] Consider the extreme case of strict, two-way

---

[13] 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 deliver meaningful interior optima that mirror observed policy choices and so have had little impact on policy.

[14] Recall that in the baseline model the ratio D/π does not affect the optimal strength of protection. Of course, to the extent that the tragedy of the anti-commons and the problem of Cournot complements raise D and reduce π for any given fixed V, they lower the equilibrium R&D expenditures and hence social welfare. However, while life is certainly more difficult when several complementary patents

complementarity in the sense that each innovation has zero stand-alone value and all must be obtained in order to operate a new technology. The analysis of this case is instructive enough that I present it in some details (see Box 3). The main conclusion is that with complementary innovations the benchmark level of protection is larger than in the baseline model. In fact, *each individual innovator* should get a share of the hypothetical profit associated with infinitely lived, complete monopoly control over the use of the new technology equal to the elasticity of supply of inventions.

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 the probability of achieving the second, where the two events are statistically independent. To focus on the problem of complementary innovations in its purest form, I assume that complementarity is strict and "two-way," in the sense that each innovation has zero stand-alone value and both must be obtained in order to operate a new technology. Therefore, the new technology can be operated, and the social and private returns from it can be netted, only with probability $x_1 x_2$. I also assume that a firm can race for innovation 1 or innovation 2, but not both. This guarantees that intellectual property rights will inevitably be 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 equally split between the two patentees. With free entry in the race for each innovation, the zero-profit condition then becomes:

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

since a firm that achieves innovation *i* now obtains a positive payoff, i.e. ½ v^P, only if innovation *j* is also achieved. Note that firms racing to obtain innovation *i* do not internalize the positive externality they exert on firms racing to obtain innovation $j \neq i$. As a consequence, there is always a no-investment equilibrium in which all firms stay 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 regularity conditions, however, there is also a stable, symmetric, free-entry equilibrium with positive investments in research. Assume that firms manage to coordinate on the equilibrium with positive investments if there is

must be assembled together in order to operate a new technology, the optimal strength of patent protection is not necessarily lower.

one (otherwise, no problem of optimal protection arises). In such a symmetric equilibrium, the zero-profit conditions reduce to

$$\frac{1}{2}x^2\beta z\frac{\pi}{r}-\alpha c(x)=0\,.$$

The social welfare function is now

$$\begin{aligned}W&=x^2\left[z\frac{\pi+(1-\beta)D}{r}+(1-z)\frac{(\pi+D)}{r}\right]-2\alpha c(x)\\&=v^Sx^2(1-\beta z)\end{aligned}.$$

where I have used the zero-profit conditions to conclude once again that society obtains a positive net benefit from the innovation only to the extent that patent protection is limited, i.e. $\beta z < 1$. The reason is that the patentees' profits are entirely dissipated in the patent races, and here I have returned to the assumption that $CS = 0$. The only difference with equation (2) is that the probability that the new technology can be operated is now x^2, because now two innovations must be achieved. Technical Appendix C shows that the optimal patent strength is now given by

(7) $\beta z = 2\eta$

i.e., *ceteris paribus* it is twice as large as in the single innovation case. This result directly extends to the case of n complementary innovations: the benchmark strength of protection is now $n\eta$. (For a free-entry equilibrium with positive investments to exists, a necessary condition is $\eta < \frac{1}{n}$, which implies that the benchmark is necessarily lower than one.)

The logic underlying this result is very simple: with strictly complementary innovations, each firm investing to obtain innovation i exerts a positive externality on firms racing to achieve innovation j. This positive externality is an additional source of distortion that tends to reduce the market equilibrium investment in research. As a consequence, it is socially desirable to raise the innovators' reward as compared to the stand-alone case.

Equation (7) has been obtained under an assumption of symmetry. Moving beyond the symmetric case, there arises a problem of appropriately dividing the profits from the invention between the inventors that concurred to the discovery (in the symmetric case, an equal split is generally optimal). For instance, with strict, two-way complementarity all components of an innovation are by definition equally valuable, but the R&D cost functions may differ across different technological components. More generally, the assumption of strict, two-way complementarity is restrictive. In these cases, the solution to the problem of the optimal division of profit may be far from obvious, and the policy tools policymakers can use to control the division of profits are limited. As a consequence, certain inventors may get more than their appropriate share and others less.

However, this paper focuses on the issue of the optimal size of the pie and is not concerned with its optimal division. In Technical Appendix C, I show that when the elasticity η is constant these two issues are indeed orthogonal: the modified elasticity rule (7) holds independently of the way total profits are divided between the various patentees. A sub-optimal split (e.g., different from fifty-fifty in the symmetric case) changes the investments in research reducing the probability of success and hence social welfare, but does not affect the benchmark strength of protection. (With a variable elasticity things are different, since changes in R&D expenditures may impact the elasticity η.)

5.5. Sequential innovation

Sequential innovation is perhaps the most important source of complementary innovations. Consider a basic innovation and a follow-on innovation that cannot be sought, and perhaps cannot even be conceived of, until the basic innovation is achieved. In this case, two innovations are actually needed for the second-generation technology to be operated: the basic innovation and the follow-on innovation. If the basic innovation has zero stand-alone value, i.e. it is a pure research tool, we are in the case of strict, two-way complementarity. The only difference with the analysis of the preceding subsection is that now the two innovations must be achieved in sequence. (This is not necessarily for bad, as it may facilitate firms' coordination on the equilibrium with positive investments in research, helping to overcome a problem discussed in Box 3.)

Things are different if instead the basic innovation has a positive stand-alone value. In this case, complementarity is "one-way": the basic innovation can now be practiced even in the absence of the follow-on innovation, but follow-on research cannot start until after the basic innovation is achieved. This means that the two innovations are now inherently asymmetric, and so the problem of the optimal division of the profit does not have an obvious solution even if the R&D cost function is the same across innovations. In addition, unless the original innovator is suitably protected, competition from second-generation innovators will erode the original patentee's profits. This creates subtle interactions between the policy tools that affect the size of the pie and those that affect the way the pie is divided. For instance, if the original innovator is granted forward patent protection through the leading breadth, this will simultaneously raise both the total size of the pie and the fraction accruing to the original innovator (Scotchmer, 2004).

As a consequence, it is likely that the modified elasticity rule (7) needs some further changes in the presence of sequential innovation. Unfortunately, so far the literature on sequential innovation has focused almost exclusively on the issue of the division of profits between subsequent generations of innovators and so little is known on the direction and the size of these changes. This is an important issue for future research. All we can say for now is that with sequential innovations the benchmark level of patent protection is greater than in the baseline model because of the dynamic complementarity between subsequent generations of innovations.

5.6. Other protection mechanisms

Surveys of US firms have repeatedly found that secrecy, lead time, and the control of complementary assets are more highly ranked than patents as a protection mechanism for both product and process innovations, and have increased in importance over the last decade (Levin *et al*, 1987; Cohen *et al*, 2000).[15] 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 a half of patentable innovations are actually patented (Duguet and Kabla, 1998). For those innovations that are patented, the additional returns accruing to innovators thanks to patent protection (i.e., the so-called patent premium) have been estimated to be equivalent to an implicit subsidy on R&D expenditure around 15-25% (see Schankerman, 1998, McCalman, 2001, and Arora *et al*, 2005). Thus, the incentivizing effect of patents seems relatively small, and certainly firms would have a substantial incentive to innovate even in the absence of patents.

This means that the assumption that in the absence of patent protection (i.e., when $\beta z = 0$) the innovator makes zero profit is evidently false. How does the elasticity rule change when this unrealistic assumption is relaxed? Here I shall argue that taking other protection mechanisms into account does not necessarily lower the benchmark obtained from the baseline model. The key insight is very simple: with constant returns to scale, secrecy, lead time and other protection devices can only provide a reward to innovators to the extent that they allow innovators to price above marginal cost.[16] But supra-competitive pricing entails deadweight losses, irrespective

---

[15] Cohen *et al* (2000) report that for process innovations, only 23% of all respondents consider patents as an effective appropriating mechanism as compared to 50% and 38% of respondents 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%).

[16] With indivisibilities or increasing marginal costs things can be different. As 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, except perhaps in rather special cases.

of the source of market power. It follows that other protection 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 where innovators can cumulate the benefits from patent protection with those of other appropriating devices and the case where other protection mechanisms are alternative to patent protection. I start from this latter case and to fix ideas I 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.) If the innovator opts for secrecy, the invention can be concealed for a time period the expected duration of which varies. The innovator will opt for patenting if and only if his expected profits under secrecy are lower than those guaranteed by patent protection. Assume for simplicity that the breadth of protection with secrecy is the same as with patents (however, the argument readily extends to the case where the breadth of protection differs). Then, those inventions for which the expected duration of the secret exceeds the patent term are kept secret, and the others are patented.

Now consider how a change (say, an increase) in the strength of patent protection impacts social welfare. By making patents relatively more attractive, such a policy move will induce some inventors to switch from secrecy to patent protection. Therefore, inventions can now be divided into three groups: those that are kept secret both before and after the move, those that are patented both before and after the move, and those that switch from secrecy to patenting. Note that the patent premium is negative for the first group, positive for the second, and zero for the third. I shall call the latter inventions “marginal”. A change in patent strength has no effect on the first group of inventions. As for the second group, comprising those inventions that would have been patented anyway, the effect is the same as if secrecy was not a feasible option. Therefore, when secrecy is an alternative to patent protection, the optimal patent strength would be the same as in the baseline model but for marginal innovations. More precisely, whether the option of keeping the innovation secret raises or lowers the optimal strength of protection depends on whether society gains or loses when marginal innovators switch from secrecy to patent protection.

If the only social costs associated with secrecy was the deadweight loss due to supra-competitive pricing, whether marginal innovations are protected by patents or secrecy would be a matter of indifference for society. (For marginal innovations, the expected duration of the secret equals the life of the patent and so the deadweight loss is the same.) In this simple case, the benchmark level of protection is unaffected by the fact that secrecy is a feasible option. However, secrecy may entail additional social costs that make a switch from secrets to patents socially desirable. First, secrecy may impede follow-on research (with patents, innovative technological knowledge must be disclosed, allowing others to conduct follow-on research). Second, secrecy does not preclude independent creation by others and so invites investments to duplicate the innovation. If such duplicative efforts are actually exerted, they add to the social costs of secrecy, making the benchmark strength of protection greater than in the baseline model.[17]

Consider next the case where innovators can cumulate the benefits from patent protection with those of other protection mechanisms. As a first approximation, one can deal with this case by simply re-interpreting the profit ratio that enters the elasticity rule (3) as an all-inclusive ratio that accounts for all profits obtained by innovators thanks to patents or other mechanisms. If the all-inclusive profit ratio is lower than the elasticity of the supply of inventions, innovators are still under-rewarded and so it would be desirable to raise the level of patent protection.

However, this broader interpretation of the profit ratio may call for some changes in the previous calculations. In particular, other protection mechanisms can raise the effective length of protection as compared to the statutory patent life. For example, brand loyalty or any other barrier to entry may allow patentees to sell their products at a premium even after the patent expires. When the inventor continues to reap positive profits after the patent expires, the benchmark level of protection is lower than in the baseline model. More specifically, suppose that a share γ of a fully protected inventor's flow profits π can be obtained forever, even after the patent expires. The appropriate elasticity rule is now

$$(8) \qquad \beta z + \gamma(1 - z) = \eta$$

where the right hand-side of (8) is the new profit ratio.

5.7. Summary

In this section I have argued that allowing for imperfect competition in research and for the possibility that society enjoys a positive net benefit from the innovation even before the patent expires unambiguously raises the benchmark strength of protection as compared to the baseline model. I have also argued that the existence of other protection mechanisms does not necessarily reduce the benchmark strength of

[17] If, however, the threat of duplication induces the inventor to pre-emptively license his trade secret, and such licensing agreements allow the inventor to appropriate all the saved duplication costs, then secrets can reward innovative activity more efficiently than patents (see Maurer and Scotchmer, 2002; Cugno and Ottoz, 2006). In this case, the optimal strength of protection would be lower in the presence of secrecy. Therefore, the issue of whether optimal patent strength is greater or lower with duplication boils down to an assessment of the prevalence and the efficiency of trade secret licensing. The special difficulty of selling unprotected ideas is well understood since Arrow (1962). The problem is particularly severe when the trade secret information is licensed to multiple parties. In this case, it is often exceedingly difficult to craft licenses permitting the licensors to maintain full control over the trade secret information. For example, some licensees might in turn secretly sell the trade secret information to others. The empirical evidence seems to confirm that trade secret licensing is not very common, and much less common than patent licensing. For example, Arora and Ceccagnoli (2006) find that in their data "only 12% of the nonpatentees license, whereas about 40% of the patentees license." They conclude that "the presence of a patent is almost essential for licensing" (p. 294).

protection. Finally, I have argued that with complementary or sequential innovation the benchmark strength of protection increases.

Overall, the extensions analyzed in this section call for an upward adjustment in the benchmark level of protection as compared to the baseline model. Such an adjustment might be substantial, although it is difficult to quantify in the absence of quantitative estimates of the additional parameters.

6. TRANSACTION COSTS, BUSINESS STEALING, AND HOLD UP

So far, I have abstracted from the administrative, enforcement and litigation costs entailed by the patent system. I have also assumed that the patentee can make a profit out of his patent only to the extent that his innovation creates new social value. This latter assumption may fail because of a business stealing effect, because agents seek patent protection opportunistically, and because patents may be used strategically to hold up firms that have infringed inadvertently.

In the presence of transaction costs, business stealing, and hold up, patents involve additional social costs beyond monopoly deadweight losses D. This section discusses these additional costs. Some of these costs are fixed, i.e. they must be borne by society just because of the existence of the patent system and independently of the strength of protection. Such fixed costs do not affect the benchmark strength of protection (although they count to determine whether society should use patents at all). Some of the costs, however, are "variable" in that their size depends on the strength of protection; as such, they can affect the optimal strength of protection.

6.1. Transaction costs

Intellectual property rights, like most legal institutions, entail a variety of transaction costs. To analyze the impact of transaction costs on the benchmark strength of protection, I distinguish between transaction costs borne by patent holders and those borne by somebody else. Although both types of costs reduce social welfare, their effects on the benchmark strength of protection are different because only transaction costs borne by patent holders reduce the incentive to invest in research.

Let TC denote flow transaction costs, and define $\psi =: \dfrac{TC}{\pi + D}$ as the ratio of transaction costs to the social value of the innovation. Let λ denote the fraction of transaction costs borne by the patent holder. Then the following modified elasticity rule obtains (see again Technical Appendix C for details):

(9) $$\beta z = \frac{\eta}{1 + (1 - \lambda)\psi}$$

Clearly, the existence of transaction costs reduces the benchmark strength of protection. However, it does so only to the extent that transaction costs do not fall on the patent holders. To be sure, even transaction costs borne by patent holders are real social costs that reduce the value of innovations and affect social welfare negatively. But they also reduce the private incentive to innovate, and therefore do not call for a downward adjustment in the strength of protection (however, they may have an indirect impact on the benchmark level of protection *via* the elasticity η).

I now discuss various types of transaction costs, trying to assess both their size and the extent to which they are borne by patent holders and thus concur to modify the elasticity rule. First, patents entail administrative costs. The patent application and review process is costly both for applicants and patent offices. However, to the extent that patent offices are funded through patent fees and patent-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 total administrative costs are financed out of general fiscal revenue. These costs probably represent a very small share of the social value of innovations.

Second, patents entail legal costs. Patent litigation is notoriously very costly. The American Intellectual Property Law Association has estimated the average costs of patent litigation at about one million $ for each party. For patents valued more than $25 millions, the average cost is $4 millions for each party (Hoti *et al*, 2006). These figures are striking, but taking into account that $25 millions is the *lower 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 lower than a 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). Moreover, most cases are settled before trial. Lemley and Shapiro (2006) report that less than 5% of litigated patents are litigated all the way to a decision, and litigation that results in settlement is probably much less costly. On the basis of this data, one can conclude that litigation costs are probably small as a share of the value of innovations in the US. And it seems unlikely that litigation costs are larger elsewhere. Moreover, around a half of total litigation costs are borne by patent holders.

Patents can entail other enforcement costs that are more difficult to measure. Most of these costs are probably associated with the prevention and detection of potential infringements, and as such are borne by patentees. However, non-patentees may also bear some costs, e.g. to obtain technical and legal advice so as not to infringe existing patents. In the US, there may be further costs to avoid so-called wilful infringement. Finally, there are transaction costs associated with patent licensing, part of which must be borne by licensees. These licensing costs can be significant, especially if intellectual property rights are highly fragmented. However, there is no quantitative evidence on their size.

6.2. Business stealing and hold up

In the baseline model I have assumed that the patentee can make a profit out of his patents only to the extent that the innovation creates new social value. This assumption may fail for various reasons. First, lured by the prospect of monopoly rents, opportunistic agents may try to seek patent protection even if they have not really innovated, or have achieved only obvious technical improvements. Overloaded patent offices may grant patents improperly, especially in technological sectors where it is difficult to ascertain what is and what is not in the prior art (Jaffe and Lerner, 2004). Yet, improperly granted patents could be easily pruned by an efficient post-grant challenge system. However, for a variety of reasons, like free riding or collusion, the system of post-grant challenges may be ineffective (Farrell and Shapiro, 2006). As a consequence, some "weak" patents may survive and be enforced.[18]

Another reason why the patentee's reward may exceed what he contributed to society is 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 the rents previously earned by those incumbents. For example, a share of the rents accruing to the author of a new economics textbook may be taken away from authors of competing textbooks (although some rents will typically come from the net increase in social welfare due to the enlarged variety of available products).

To analyze the impact of these phenomena on the optimal strength of protection, I assume that only a fraction 1-ϕ of the innovator's flow profits π corresponds to net social value. The parameter ϕ can be thought of as the coefficient of business stealing, i.e., the fraction of the innovator's rents that are stolen from previous incumbents. Alternatively, ϕ can be interpreted as the share of bad patents, i.e. the probability that a patent is granted in the absence of a genuine innovation. The modified elasticity rule then becomes:

$$(10) \qquad \beta z = \frac{(1-\phi)\pi + D}{\pi + D}\eta$$

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

Yet another reason why a patentee's profit can exceed his contribution to social welfare is that a patentee can use his patents to hold up manufacturers that have infringed inadvertently. Shapiro (2006) argues that the design of innovative products may exhibit a putty-clay pattern: *ex ante*, the product can be easily designed in various alternative ways, but *ex post* it may be very difficult and time consuming to modify the product's design. With complex technologies characterized by fragmented

[18] On the other hand, the patent system may also err in the opposite direction, i.e. by not granting a patent for genuine inventions. If the parameter β also captures the probability that a genuine invention is denied patent protection, this effect should be accounted for when assessing patent breadth.

intellectual property rights, it can happen that the manufacturer inadvertently infringes on some patents that *ex ante* he could have circumvented easily (this differentiates this scenario from the strict, two-way complementarity case analyzed in section 5.4). *Ex post*, the patent holder acquires a strong, undeserved bargaining power that may allow him to extract large rents from the manufacturer. Shapiro argues that in this scenario patent holders are systematically over-compensated. Formally, this hold up problem has the same effects as business stealing, i.e. it allows the patent holder to obtain a share of the rents that otherwise would accrue to somebody else.

6.3. Summary

Transaction costs, business stealing, dubious patents, and hold up are all phenomena that call for a downward adjustment in the benchmark strength of protection as compared to the baseline model. While the importance of these phenomena is undeniable, the size of these effects is largely unknown. As for transaction costs, I have argued that they are probably small as a share of the social value of innovations. As for the other effects, the available evidence is mainly anecdotal.

To some extent, the size of these effects is a matter of policy. For example, the phenomenon of business stealing is typically associated with sequential innovation, and its importance negatively depends on the strength of forward patent protection guaranteed by the novelty requirement and leading breadth (Scotchmer, 2004). The hold up problem could be addressed by appropriately calculating damages in case of infringement. The risk of granting dubious patents could be reduced by increasing the resources devoted to patent offices, by appropriately changing their incentives, by facilitating post-grant oppositions etc.

However, even if these problems can be somewhat alleviated, one must recognize that they are here to stay and so they must be taken into account when determining the benchmark strength of protection. The fact that it is difficult to empirically assess their importance suggests extra caution in drawing policy conclusions.

7. CONCLUSION

There is strong empirical evidence that if it was possible to raise R&D investment by fiat, actual R&D investment should be at least doubled (Jones and Williams, 1998). Unfortunately, we live in a second best world where it is not possible to raise R&D investment by fiat and even highly imperfect policy tools like patents can be useful. This paper asks whether in such a second-best world we have gone too far in using patents to incentivize innovative activity; in other words, to repeat the title of the paper, whether patents over-compensate innovators. To address this issue, I have developed a simple and flexible model of the optimal strength of patent protection. In particular, my baseline model delivers a simple elasticity rule that asserts that innovators are over-compensated if the profit ratio exceeds the elasticity of the supply of inventions. I have then tried to combine this simple result and available empirical evidence to assess the over-compensation hypothesis.

The main conclusion of this exercise is that the over-compensation hypothesis does not seem to be supported by the data. It must be stressed that this conclusion is very preliminary and tentative. Estimates of the elasticity of the supply of inventions vary considerably. As for the profit ratio, my rudimentary attempts at measurement resulted in what at best might be considered an educated guess. In addition, there is a desperate lack of hard evidence concerning some of the effects analyzed in section 6, like business stealing, dubious patents and hold up, which might overturn the conclusions reached on the basis of the baseline model. More empirical evidence is therefore needed to provide a more precise assessment of the over-compensation hypothesis. But a preponderance of what evidence is currently available suggests that patentees are not over-compensated.

Although future research might overturn this specific conclusion, the approach developed in this paper may nonetheless prove useful. The simplicity and flexibility of the elasticity rule might convince other researchers that a more accurate assessment of the over or under-reward hypotheses is not out of reach, and help clarify what kind of additional empirical research is most needed to provide more reliable answers.

References

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*, Princeton, N.J., Princeton University Press, 609-625.

Bester, H. and E. Petrakis, (2003), 'Wages and Productivity Growth in a Competitive Industry', *Journal of Economic Theory*, 109, 52-69.

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.

Boldrin, M. and D. Levine (2005), 'IP and Market Size,' mimeo, Department of Economics, UCLA.

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.

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 Not),' 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.

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. (1990), 'Patent Policy and Costly Imitation', *RAND Journal of Economics*, 23(1), 52-63.

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.

Grossman, G.M. and E. L.C. Lai (2004), ‘International Protection of Intellectual Property,’ *American Economic Review*, 94(5), 1635-1653.

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.

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.

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.
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, 1-39.

Jaffe, A.B. (2000) ‘The U.S. Patent System in Transition: Policy Innovation and the Innovation Process,’ *Research Policy* 29 (4–5), 531–557.

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.

Kortum, S. and J. Lerner (1998), ‘Stronger Protection or Technological Revolution: Which is Behind the Recent Surge in Patenting?’ *Carnegie-Rochester Conference Series on Public Policy*, 48, pp. 247-304.

Landes, W. and R. Posner (2003), *The Economic Structure of Intellectual Property Law*, Harvard, Harvard University Press.

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,' *The Journal of Law and Economics*, 45, 643–672.

Linn, J. and D. Acemoglu (2004) 'Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry,' *Quarterly Journal of Economics*, 119(3), 1049-1090.

Mansfield, E., Schwartz M. and S. Wagner (1981), 'Imitation Costs and Patents: An Empirical Study,' *Economic Journal*, 91, 907-918.

Maurer, S. P. and S. Scotchmer (2002), 'The Independent Invention Defense in Intellectual Property,' *Economica*, 69, 535-547.

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.

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.

Priest, G. (1986), 'What Economists Can Tell Lawyers about Intellectual Property: Comments on Cheung,' in Palmer, J. and R. Zerba (eds.), *Research in Law and Economics*, London, JAI Press, 20-1.

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. (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.

Shapiro, C. and M. A. Lemley (2006), 'Patent Hold-Up and Royalty Staking,' mimeo, University of California Berkeley.

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.

Technical Appendix A: Alternative models of innovation

In this Appendix I show that equation (1) can be regarded as a reduced form of various models of innovation that are commonly employed is the economics literature. More precisely, for the first two models developed below the equivalence holds for both the free-entry and the monopoly equilibrium. The most natural interpretation of the last two examples, however, refers to the case of monopoly in research.

A1. Poisson model

Consider 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:

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

where y_i is i's R&D effort and also i's instantaneous probability of innovating (conditional on no previous success), $Y = \sum_{j=1}^{m} y_j$ is the aggregate R&D effort, m is the number of firms in the research industry, r is the discount rate, e^{-Yt} is the probability that no firm has innovated yet by time t, and αy_i is the R&D cost.

Aggregate profits are $\frac{Y}{Y+r} v^P - \alpha Y$, so that $x := \frac{Y}{Y+r}$ can be thought of as the "discounting adjusted" probability of success. The idea here is that 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 is $c(x) = \frac{rx}{1-x}$. Note that it is increasing, convex, and satisfies $c(0) = 0$.

With free entry in the R&D industry, the zero-profit condition (1) fully determines the aggregate R&D effort Y: $Y = \max\left[\frac{v^P}{\alpha} - r, 0\right]$ or equivalently $x = \max\left[1 - \frac{r\alpha}{v^P}, 0\right]$. Note that with constant returns in research, the equilibrium number of active firms is indeterminate and only aggregate R&D investment is determined. However, the model readily extends to the case of variable returns. For example, if each firm active in research must pay a fixed cost and the research technology exhibits increasing marginal costs, the equilibrium number of active firms is fully determined. The model can also be used to analyze the case of monopoly in research, and the case of competition in research 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 $v^P e^{-rt} - \alpha\tilde{c}(t)$, where $\alpha\tilde{c}(t)$ is the discounted cost of innovating at time *t*. Assume that $\lim_{t\to 0} \tilde{c}(t) = \infty$ and that $\tilde{c}(t)$ decreases rapidly enough with *t*. Define $x := e^{-rt}$, such that $t = \frac{\log x}{-r}$. The profit function then becomes $xv^P - \alpha c(x)$ where $c(x) := \tilde{c}\left(\frac{\log x}{-r}\right)$. 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 $xv^P - \alpha c(x)$. Thus, this model is formally identical to the model used in the paper.

A3. Fixed, uncertain R&D costs

Here I follow Scotchmer (2004). Suppose that a research firm has 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 cost is then $c(x) = \int_0^{G^{-1}(x)} cg(c)dc$. In particular, the firm will invest if and only if $\alpha c \leq v^P$. Therefore,

the *ex ante* probability of success is $G\left(\frac{v^P}{\alpha}\right)$. But this is equivalent to assuming that the research firm maximizes its expected profits $xv^P - \alpha c(x)$, as can be confirmed by checking that the first-order condition of this maximization problem is
$v^P = \alpha \frac{dG^{-1}(x)}{dx} G^{-1}(x) g\left[G^{-1}(x)\right]$, or equivalently $v^P = \alpha G^{-1}(x)$. In other words, the optimal cut-off is indeed $\frac{v^P}{\alpha}$.

A4. Variable quality

As a final example, consider a firm that can create a new product of variable quality. Let x denote the quality of the product, and let v^P denote the private value 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.

Additional references for Technical Appendix A

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

Loury, G. C. (1979), 'Market Structure and Innovation,' *Quarterly Journal of Economics*, August, 395-410.

Technical Appendix B: Derivation of the elasticity rule

Note first of all that both x and W depend only on the product βz, i.e. on the strength of protection. Differentiating (2) with respect to βz one gets the following first order condition for a maximum:

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

I assume for simplicity that the parameter α is such that the social problem entails an interior solution. Using the zero-profit condition (1), by the implicit function theorem one gets

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

Substituting into (B1) and rearranging one obtains

(B3) $$\frac{dW}{d(\beta z)} = \frac{\pi + D}{r}x\left[\frac{\eta}{1-\eta}\frac{1-\beta z}{\beta z} - 1\right] = 0$$

whence equation (3) follows immediately.

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

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

These conditions determine the equilibrium level of investment in research as a function of innovation type s, i.e. $x(s)$. The social welfare function becomes

(B5)
$$W = \int \left[xz\frac{\pi(s) + (1-\beta)D(s)}{r} + x(1-z)\frac{\pi(s) + D(s)}{r} - \alpha c(x,s) \right] f(s)ds$$
$$= \int \frac{\pi(s) + D(s)}{r} x(s)(1-\beta z) f(s)ds$$

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

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

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

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

The right hand side of (B7) can be interpreted as a weighted average of the individual elasticities $\eta(s)$. The weights $\frac{[\pi(s) + D(s)]}{1-\eta(s)} x(s) f(s)$ reflect the value of innovations and their relative frequency. They are also increasing in the elasticity $\eta(s)$, implying that a mean preserving spread in the distribution of the elasticities $\eta(s)$ raises the benchmark strength of protection.

Technical Appendix C: Derivation of the modified elasticity rules

C1. Drastic innovations and spillovers

With $CS > 0$, the zero-profit condition (1) does not change but the social welfare function becomes

(C1)
$$W = x\left[z\frac{\pi + (1-\beta)D + CS}{r} + (1-z)\frac{(\pi + D + CS)}{r}\right] - \alpha c(x)$$
$$= \frac{(\pi + D)}{r}x(1-\beta z) + x\frac{CS}{r}$$

The first order condition for a maximum becomes

(C2)
$$\frac{\pi + D}{r}\left[\frac{dx}{d(\beta z)}(1-\beta z + \sigma) - x\right] = 0$$

From this equation, using (B2) one obtains

(C3)
$$\frac{\pi + D}{r}\left[\frac{\eta}{1-\eta}(1-\beta z + \sigma)\frac{x}{\beta z} - x\right] = 0$$

whence the modified elasticity rule (4) easily follows.

C2. Monopoly in research

For simplicity, in this subsection I assume that the R&D expenditure function $c(x)$ is iso-elastic: $c(x) = x^{\frac{1}{\eta}}$. With monopoly in research, the zero-profit condition (1) must be replaced by the first order condition of the monopolist's maximization problem, i.e.:

(C4)
$$\beta z\frac{\pi}{r} - \frac{\alpha}{\eta}x^{\frac{1}{\eta}-1} = 0$$

It can be shown that equation (B2) continues to hold, but the social welfare function is now given by the first line of equation (2) (the second line follows only if the zero-profit condition holds). Using (C4), social welfare now simplifies to:

(C5) $$W = \frac{(\pi + D)}{r} x(1 - \beta z) - \alpha x^{\frac{1}{\eta}} \left(1 - \frac{1}{\eta}\right)$$

Differentiating (C5) one obtains the following first order condition for a maximum:

(C6) $$\frac{dx}{d(\beta z)} \left[\frac{(\pi + D)}{r} (1 - \beta z) - \frac{\pi}{r} \beta z \left(1 - \frac{1}{\eta}\right) \right] = 0$$

where I have used the first order condition (C4) to simplify the second term inside square brackets. Using (B2) and rearranging terms one easily obtains the modified elasticity rule (5).

C3. The optimal combination of breadth and length

Following the discussion is section 5.3, assume that the deadweight loss is an increasing function of the breadth of protection, $D(\beta)$, with $D(1) = D$ and $D(0) = 0$. Social welfare then becomes

(C7) $$W = x\left[z\frac{\pi + D - D(\beta)}{r} + (1 - z)\frac{(\pi + D)}{r} \right] - \alpha c(x)$$

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 $D''(\beta) > 0$. Since the optimal policy will then have $z = 1$, taking into account the zero-profit condition social welfare reduces to

(C8) $$W = x\left[\frac{(1 - \beta)\pi}{r} + \frac{D - D(\beta)}{r} \right]$$

The first-order condition for a maximum becomes

(C9) $$\frac{dx}{d\beta}\left[\frac{(1 - \beta)\pi}{r} + \frac{D - D(\beta)}{r} \right] - x\frac{\pi + D'(\beta)}{r} = 0$$

Simplifying and rearranging one gets the following modified elasticity rule

(C10) $$\beta = \frac{\eta}{\Xi + \eta(1 - \Xi)}$$

where $\Xi := \dfrac{\pi + D'(\beta)}{\pi + \dfrac{D - D(\beta)}{1-\beta}}$. Although (C10) does not provide a closed-form solution for *β*, note that $D''(\beta) > 0$ implies $D'(\beta) < \dfrac{D - D(\beta)}{1-\beta}$. 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.

C4. Complementary innovations

Consider the case of two strictly complementary innovations analyzed in Box 3. The extension to the general case of *n* complementary innovations is straightforward. As explained in Box 3, the zero profit conditions are:

(C11) $$\frac{1}{2}x^2\beta z\frac{\pi}{r} - \alpha c(x) = 0 .$$

The social welfare function is now

(C12) $$\begin{aligned} W &= x^2\left[z\frac{\pi + (1-\beta)D}{r} + (1-z)\frac{(\pi + D)}{r}\right] - 2\alpha c(x) \\ &= \frac{(\pi + D)}{r}x^2(1-\beta z) \end{aligned} .$$

From the zero-profit condition one now obtains:

(C13) $$\frac{dx}{d(\beta z)} = \frac{\eta}{1-2\eta}\frac{x}{\beta z}$$

Differentiating (C12) and using (C13) one obtains

(C14) $$\frac{dW}{d(\beta z)} = \frac{\pi + D}{r}x^2\left[\frac{2\eta}{1-2\eta}\frac{1-\beta z}{\beta z} - 1\right] = 0$$

whence the modified elasticity rule (7) immediately follows.

To analyze the case of an arbitrary division of total profit *π* between the two patentees, suppose that the first patentee gets a share *γ* of the profits and the second patentee the remaining share (1-γ). Assume that the R&D expenditure function $c(x)$ is iso-elastic: $c(x) = x^{\frac{1}{\eta}}$. The two zero-profit conditions become

(C15) $$\gamma x_1 x_2 v^P - \alpha_1 x_1^{\frac{1}{\eta}} = 0$$

and

(C16) $$(1-\gamma) x_1 x_2 v^P - \alpha_2 x_2^{\frac{1}{\eta}} = 0$$

respectively. From (C15) and (C16) one gets

(C17) $$x_1 = \left[\frac{(1-\gamma)\alpha_2}{\gamma\alpha_1}\right]^{\eta} x_2$$

which can be substituted back into (C16) to get

(C18) $$(1-\gamma)\left[\frac{(1-\gamma)\alpha_2}{\gamma\alpha_1}\right]^{\eta} v^P = \alpha_2 x_2^{\frac{1}{\eta}-2}$$

From (C18) and (C17) one gets

(C19) $$\frac{dx_2}{d(\beta z)} = \frac{\eta}{1-\eta}\frac{x_2}{\beta z}$$

The social welfare function becomes

(C20) $$W = \frac{(\pi + D)}{r}\left[\frac{(1-\gamma)\alpha_2}{\gamma\alpha_1}\right]^{\eta} x_2^2 (1-\beta z)$$

and the modified elasticity rule (7) follows exactly like in the symmetric case.

C5. Other protection mechanisms

Suppose that a share γ of a fully protected inventor's flow profits π can be obtained even after the patent expires. The zero-profit condition then becomes

(C21) $$x\left[\frac{z\beta\pi}{r} + (1-z)\frac{\gamma\pi}{r}\right] - \alpha c(x) = 0$$

and the social welfare function becomes

(C22)
$$W = x\left[z\frac{\pi + (1-\beta)D}{r} + (1-z)\frac{(\pi + (1-\gamma)D)}{r}\right] - \alpha c(x)$$
$$= \frac{\pi + D}{r}x\left[1 - \beta z - \gamma(1-z)\right]$$

since a share γ of the deadweight loss D is now borne forever. In this framework, the profit ratio is $\beta z + \gamma(1-z)$, and it is this profit ratio that must be set equal to the elasticity η. This gives us immediately the modified elasticity rule (8).

C6. Transaction costs

Assume that transaction costs are variable in the sense that they are proportional to the strength of protection. In the presence of transaction costs the zero-profit condition becomes

(C23)
$$x\left[\frac{z\beta\pi}{r} - \lambda z\beta\frac{TC}{r}\right] - \alpha c(x) = 0$$

where λ is the share of transaction costs borne by the patentee, and the social welfare function becomes

(C24)
$$W = x\left[z\frac{\pi + (1-\beta)D - \beta TC}{r} + (1-z)\frac{\pi + D}{r}\right] - \alpha c(x)$$
$$= x\left[\frac{\pi + D}{r}(1-\beta z) - \beta z\frac{(1-\lambda)TC}{r}\right]$$

where the second line of (C24) follows by inserting the zero-profit condition (C23) into the first line. The first order condition for a maximum becomes

(C25)
$$\frac{\pi + D}{r}\left[\frac{dx}{d(\beta z)}(1 - \beta z - \beta z(1-\lambda)\psi) - x - x(1-\lambda)\psi\right] = 0$$

By implicit differentiation of (C23), it can be shown that equation (B2) continues to hold. From (C25), using (B2) one obtains

(C26) $$\frac{\pi+D}{r}\left[\frac{\eta}{1-\eta}(1-\beta z-\beta z(1-\lambda)\psi)\frac{x}{\beta z}-x-x(1-\lambda)\psi\right]=0$$

Rearranging and simplifying, the modified elasticity rule (9) easily follows.

C7. Business stealing

If a share ϕ of the patentee's profits does not correspond to true social value, the social welfare function becomes

(C27) $$W=x\left[z\frac{(1-\phi)\pi+(1-\beta)D}{r}+(1-z)\frac{(1-\phi)\pi+D}{r}\right]-\alpha c(x)$$

In fact, with business stealing social welfare should perhaps include the reduced deadweight losses on "old" goods, since business stealing is usually associated with more intense competition in the product market. On the other hand, with bad patents the deadweight losses associated with improperly granted monopoly power should be deducted. Although it would be easy to account for these complications explicitly, here for simplicity I abstract from them.

Using the zero-profit condition (1), the social welfare function (C27) now simplifies to

(C28) $$W=\frac{(\pi+D)}{r}x(1-\beta z)-x\phi\frac{\pi}{r}$$

The first order condition for a maximum becomes

(C29) $$\frac{\pi+D}{r}\left[\frac{dx}{d(\beta z)}(1-\beta z-\phi\frac{\pi}{\pi+D})-x\right]=0$$

From this equation, using (B2) one obtains

(C30) $$\frac{\pi+D}{r}\left[\frac{\eta}{1-\eta}\left(1-\beta z-\phi\frac{\pi}{\pi+D}\right)\frac{x}{\beta z}-x\right]=0$$

Simple algebra then suffices to obtain the modified elasticity rule with business stealing (10).