GROWTH AND INTELLECTUAL PROPERTY

MICHELE BOLDRIN AND DAVID K. LEVINE

University of Minnesota, the Federal Reserve Bank of Minneapolis and CEPR; UCLA and the Federal Reserve Bank of Minneapolis

ABSTRACT. Intellectual property (IP) protection involves a trade-off between the undesirability of monopoly and the desirable encouragement of creation and innovation. Optimal policy depends on the quantitative strength of these two forces. We give a quantitative assessment of current IP policies. We focus particularly on the scale of the market, showing that as it increases, due either to growth or to the expansion of trade, IP protection should be reduced.

Keywords: Intellectual Property, Innovation, Monopoly.

JEL Classification: X15; X16

This paper is based upon work supported by the National Science Foundation under Grants SES 01-14147 and 03-14713. Boldrin also acknowledges research support from the Spanish SEC2005-08793-C04-01. Our ideas benefited from comments at the Theory Workshop at Columbia University, the brown bag workshop at the Dallas Federal Reserve Bank, the faculty seminar at the Chinese University of Hong Kong, the Theory Workshop at ASU, the CSIP seminar at the San Francisco Federal Reserve Bank, and from discussions with Kyle Bagwell, Gene Grossman, Chad Jones, Pete Klenow, Paul Romer, Suzanne Scotchmer and Daniel Wilson. Thanks are also due to Hengjie Ai research assistance in collecting and analyzing data.

Corresponding Author: David K. Levine, Department of Economics, UCLA, Los Angeles, CA 90095, USA. Phone/Fax: 310-825-3810. Email: david@dklevine.com.

First Version: 2nd December, 2003, This Version: 1st February 2006.

1. Introduction

In this paper we study the interactions determining the optimal level of intellectual property (IP) protection. We are especially interested in how optimal protection should change as the size of the economy grows. There is a large literature that explores the qualitative aspects of optimal IP policy, and a significant empirical literature that attempts to measure such things as the value of patents. There is, however, little connection between the two. Our goal is to use a relatively standard model of IP, based on that of Grossman and Lai [2004], and examine the policy implications of existing quantitative findings. To do this we proceed in three steps; first we use the model to derive the optimal IP policy as a function of measurable parameters and the size of the market, second we use available estimates to calibrate those parameters and quantify changes in market size, finally we put everything together to obtain the implications for optimal policy.

Innovations are introduced when they become profitable. Profitability of an innovation depends upon three factors: the initial cost of discovery, the elasticity of demand, and the size of the market; all these elements vary widely and unsystematically across innovations. We focus primarily on market size as it is easier to measure than the other two and it has grown steadily and substantially since current patent and copyright legislation was first adopted.

Theory suggests that optimal policy involves a trade-off between increasing the monopolistic distortion on infra-marginal ideas, and increasing the number of usable ideas by encouraging innovating at the margin. As the scale of the market increases, depending on elasticity of the total demand for innovative goods it may be desirable to give up some of the additional marginal ideas in exchange for reduction of monopoly across the broad variety of infra-marginal ideas that will be produced anyway. In this case the optimal policy should reduce the length of protection as the scale of the market increases. Our analysis shows that this is the empirically relevant case.

We utilize a standard model in which IP protection is socially beneficial.[1] Ideas are created subject to a fixed cost and are not appropriable per-se. There are many possible ideas yielding different *private returns* – defined as the ratio of expected monopoly revenue to cost of creation in a market of unit size. We focus on the case in which the private return is proportional to the social return to an idea. This model is related to a series of papers by Grossman and Helpman [1991, 1994, 1995] studying innovation in a Dixit-Stiglitz framework. It is most closely related, however, to Grossman

[1] We have examined the shortcomings of the "standard model" in Boldrin and Levine [1999, 2002, 2004, 2005] where we argue that IP is not generally socially beneficial.

and Lai [2004]. Both they and we show that as the total monopoly revenue function has increasing (decreasing) elasticity, then optimal protection locally decreases (increases). From a theoretical perspective, their approach differs from ours in two respects. First, where we use a static analysis, they embed the static model in a dynamic setting by treating costs and profits as time-flows, and examining balanced growth paths. Since they have already provided this interpretation, we simply note that this procedure is equally valid for our model. Second, their model uses a production function approach to the creation of new ideas. That is, ideas are of homogeneous quality, and are produced using a constant returns technology with human capital and labor as inputs. We use a disaggregated model in which ideas are heterogeneous both in their quality and in their cost of creation. The latter allows us to show that the elasticity of the total monopoly revenue function depends on the distribution of private returns from ideas. It is a useful tool for quantitative analysis, as it can be estimated from available data. This, together with the characterization of the optimal policy mentioned above, is the bottom line of the theoretical section.

In the empirical part, then, our goal is to measure whether the elasticity of total monopoly revenue is increasing – implying that protection should decline with growth in market scale – or decreasing – implying the reverse. We use two methods of measurement. First, we attempt to measure directly how total monopoly revenue varies with the private return. We use data we collected on the distribution of book revenues; and we also examine existing estimates of the value of patents. Second, the elasticity of total monopoly revenue has implications for the demand for (skilled) labor: when elasticity is constant or decreasing, the demand for labor increases by more than the increase in market scale. We examine three sources of data on labor demand: a time series on copyright; a time series on patents and R&D expenditures; and a cross-section on R&D expenditure across countries. As we discuss, each of the different sources of data has different caveats associated with it – but each supports the notion that elasticity is decreasing and that protection should optimally decrease with the scale of the market.

We conclude with a policy section, in which we examine the implications of our estimates for the level and rate of change of optimal IP protection. Our basic conclusion is that the length of patent and copyright are moving in the wrong direction, and bear little relation to the social optimum.

2. The Model

Ideas are costly to produce, and yield revenues to their creators. Key to whether a particular idea will be produced is its return – the ratio of revenue to cost of production. Revenue depends on demand – in our analysis demand changes with the size of the economy: the bigger the economy, the more demand for any particular idea. Demand also depends on institutional arrangements: if the creator has a monopoly over his idea he may earn more revenue than otherwise.

Our notion of equilibrium is that of a *patent equilibrium* in which there is a fixed common length of patent protection for all ideas. This means that, in terms of present value of the flow of consumption, a fraction $0 \leq \phi \leq 1$ occurs under monopoly, and a fraction $(1-\phi)$ occurs under competition; hence ϕ is the level or the extent of protection. While the patent lasts, the innovator is a monopolist. Once a patent expires, anyone who wishes to do so may freely make copies, and price falls to marginal cost with the innovator earning no further revenue.[2] An idea is produced if, given the patent length ϕ, the prospective monopolist finds it profitable to pay the cost of invention.

The cost of innovation depends on the price of inputs used to produce the idea – in practice the relevant input is specialized labor, the cost of which is determined by the wage rate w. Suppose the size of the economy is fixed at one, that the wage rate $w = 1$ and that the creator has a complete monopoly over his creation, so that $\phi = 1$. Under these conditions, we denote by ρ the *private return* on the creation – the ratio of present value revenue to cost. If in fact the size of the economy is λ, the wage rate is w and the creator can appropriate only a fraction ϕ of the monopoly revenue, then the return on the creation is $\lambda\phi\rho/w$. Naturally, the creator will choose to create if and only if $\lambda\phi\rho/w \geq 1$, that is $\rho \geq w/\lambda\phi$.

Ideas naturally vary in their private return ρ. Some very good ideas are valuable but cost little to produce and so have high values of ρ; others will cost a great deal and not yield much revenue, so ρ will be low. As we have observed, the ideas that will be produced are those for which the private return ρ is at least $w/\lambda\phi$. What matters from an economy wide perspective, then, is the distribution of private returns. It is convenient to think of this in terms of the amount of labor needed to create ideas. In particular, we denote by $h(\rho)$ the total amount of labor input required to produce all ideas that have private return ρ in an economy of unit size ($\lambda = 1$). The function $h(\rho)$ is similar to a probability density function, except that, since it is measured in units of labor, it need not integrate to one. For example, it is useful to compute the total amount of labor needed to produce all ideas in an economy of unit size with private returns at least equal to ρ. This is

---

[2] We assume there are no competitive rents after the patent expires; as pointed out in Boldrin and Levine [1999], inventors generally do earn positive competitive rents.

simply

$$H(\rho) = \int_{\rho}^{\infty} h(\rho')d\rho',$$

which is similar to a cumulative distribution function, except in reverse. Notice however, that it may be that as $\rho \to 0$ it may be that $H(\rho) \to \infty$, that is, we do not assume it is necessarily feasible to produce all possible ideas.

The function $H(\rho)$ is a measure of the total labor input required in an economy of unit size to produce all ideas with return at least ρ. Output can be measured by the total monopoly revenue

$$M(\rho) = \int_{\rho}^{\infty} \rho' h(\rho')d\rho'.$$

In other words, $M(\rho)$ is the sum of monopoly revenue over all ideas with private value of ρ, or greater, in an economy of unit size. We assume that M is differentiable and define the *elasticity of total monopoly revenue,* with respect to variations in the marginal idea, as $\Upsilon(\rho) \equiv -\rho M'(\rho)/M(\rho) > 0$. We also make the regularity assumption that $\Upsilon(\rho)$ is differentiable.

We need also to consider what happens to the number of ideas that are produced as the economy is scaled up. That is, in the unit economy, producing all ideas with private return ρ requires $h(\rho)$ units of labor. If an economy of size λ has λ times as many ideas with private return ρ as an economy of unit size, then this economy would require $\lambda h(\rho)$ units of labor as input. More generally, we assume that an economy of size λ has $g(\lambda)$ times as many ideas with private return ρ as an economy of unit size. To capture the principle that in a larger population more ideas of a given private return are available $g(\lambda)$ is assumed non-decreasing; without loss of generality we take $g(1) = 1$.

Our primary interest is in social welfare. To analyze welfare, we need to relate the social to the private return ρ on an idea. Our basic hypothesis is that the two are proportional. That is, under monopoly, we assume that the social return is $\overline{\nu}^M\rho$, while once the monopoly expires, the social return is $\overline{\nu}^C\rho$, where $\overline{\nu}^C > \overline{\nu}^M > 1$. The per capita social welfare corresponding to a particular level of protection ϕ, when all ideas with private return of ρ and greater are produced is:

$$W = g(\lambda)\int_{\rho}[\lambda\phi\overline{\nu}^M\rho' + (1-\phi)\lambda\overline{\nu}^C\rho' - 1]h(\rho')d\rho'$$

3. Optimal IP Protection

We first ask how socially optimal protection $\hat{\phi}$ depends on market size λ.

Proposition 3.1. *Suppose that the opportunity cost of skilled labor is constant at $w = 1$. When $\hat{\phi}(\lambda) < 1$, in a neighborhood of $\rho = 1/\lambda\hat{\phi}(\lambda)$, the* *following holds. (I)* $\Upsilon'(\rho) > 0$ *implies* $\hat{\phi}(\lambda)$ *is unique and strictly decreasing; (II)* $\Upsilon'(\rho) = 0$ *implies* $\hat{\phi}(\lambda)$ *is unique and constant; and (III)* $\Upsilon'(\rho) < 0$ *and* $\hat{\phi}(\lambda)$ *unique*[3] *implies* $\hat{\phi}(\lambda)$ *is strictly increasing.*

The details of the proof are in Appendix 1.

Next we examine the implications of increasing the scale of the market on the demand for skilled labor. Continuing to hold the wage rate fixed at $w = 1$, labor demand is given by

$$L^D(\lambda) = g(\lambda)H(1/\phi\lambda),$$

from which, letting $\mathfrak{E}$ denote the elasticity operator, we have

$$\mathfrak{E}[L^D(\lambda)] = \mathfrak{E}[g(\lambda)] - \mathfrak{E}[H(\rho)].$$

Depending on which assumptions one makes about $g(\lambda)$, the first factor ranges from zero to any large positive number. For example, in the Grossman and Lai [2004] setting, $g(\lambda)$ can be identified with aggregate human capital. To the extent this is constant, $\mathfrak{E}[g(\lambda)] = 0$. In models of growth and innovation due to externalities or increasing returns, such as Grossman and Helpman [1991, 1994, 1995] or Romer [1990], $g(\lambda)$ is assumed to increase faster than λ, hence $\mathfrak{E}[g(\lambda)] > 1$. A benchmark case is that in which each individual draws her own ideas from the same urn, either with or without replacement. If sampling is without replacement, and each person draws the same number of ideas, then $g(\lambda) = \lambda$ and $\mathfrak{E}[g(\lambda)] = 1$; if sampling is with replacement then $\mathfrak{E}[g(\lambda)] \leq 1$.

As for the second factor, notice first that the demand for labor is linked to the total revenue function by the following relation

$$H(\rho) = \int_{\rho}^{\infty} -[DM(\rho')/\rho']d\rho'.$$

Now, assume that $M(\rho) = \rho^{-\zeta}$, which is the constant elasticity case. Then

$$H(\rho) = \frac{(\zeta+1)(\rho)^{-1-\zeta}}{\zeta+2}$$

and

$$\mathfrak{E}[L^D(\lambda)] = \mathfrak{E}[g(\lambda)] + \zeta + 1.$$

Notice that when $\mathfrak{E}[g(\lambda)] > 1 - \zeta$, the elasticity of labor demand is predicted to be larger than two, hence the elasticity of per capita labor demand is greater than one. More generally, since $\mathfrak{E}[g(\lambda)] \geq 0$, we have $\mathfrak{E}[L^D(\lambda)/\lambda] > 0$. In other words, as the size of the economy grows, the *share* of workers in the idea sector grows as well. The next proposition, proven in Appendix

[3]In this case we cannot guarantee that the second order condition is satisfied, so we must rule out the possibility that $\hat{\phi}(\lambda)$ has multiple values.

1, shows how to extend from the case of constant elasticity to decreasing elasticity of the total monopoly revenue.

Proposition 3.2. *Consider two aggregate monopoly revenue functions M_1, M_2 that have the same value $M_1(\rho) = M_2(\rho)$ and derivative $DM_1(\rho) = DM_2(\rho)$ (hence, $\Upsilon_1(\rho) = \Upsilon_2(\rho)$) at ρ. If $D\Upsilon_1(\rho') < D\Upsilon_2(\rho')$ for $\rho' \geq \rho$, then*

(1) *Labor demand associated to M_1 is smaller than the one associated to M_2; that is,*
$$\int_\rho^\infty -[DM_1(\rho')/\rho']d\rho' < \int_\rho^\infty -[DM_2(\rho')/\rho']d\rho'.$$
(2) *The elasticity of labor demand from M_1 is greater than the elasticity of labor demand from M_2, that is $\mathfrak{E}[H_1(\rho)] > \mathfrak{E}[H_2(\rho)]$.*
(3) *As the elasticity of total revenue goes from increasing, to constant, to decreasing, the elasticity of the associated labor demand functions increases monotonically.*

In plain words: a revenue function with decreasing elasticity implies an elasticity of labor supply even larger than that of a constant elasticity revenue function, which we have shown to be at least one in practice. Playing this backward: should the empirical elasticity of per capita labor supply with respect to market size be smaller than one, then the associated total revenue function must display increasing elasticity. Per capita labor in the idea sector growing faster than the scale of market is consistent with increasing elasticity of total monopoly revenue, because $\mathfrak{E}[g(\lambda)]$ can be large, which is independent of the elasticity of monopoly revenue. However, if per capita labor grows more slowly than the size of the market, we must rule out both constant and decreasing elasticity.

4. Empirical Analysis of Total Monopoly Revenue

Up until now we have been thinking of ideas as empty boxes to be filled in by individuals. From an empirical perspective, it is more useful to think of each individual being associated with his own ideas and his own opportunity costs of engaging in innovative activity. We then identify individuals with their private returns ρ and think of them as equivalent to the expected value of their ideas, with the latter being drawn from an underlying distribution $\mu(\rho)$ satisfying the restrictions discussed earlier. We are interested in the shape of $\mu(\rho)$ as this would allow us to compute the elasticity of $M(\rho)$ at the "cutoff idea-individual."

An issue arises at this point. In the available data we observe revenues not returns, ρ. Further, it is hard to observe directly either the opportunity cost, w, of each inventor or the labor cost of his ideas. Hence we need to assume that, for all the ideas in the data set, the cost of producing that idea is the same. This ensures that returns are proportional to revenues. If our data sets contained ideas produced in very different sectors, this assumption would be absurd. To avoid this we will try to restrict attention to sets of goods that are relatively homogeneous, so that it is reasonable to assume that the cost of creation is roughly constant within each set.

A second issue also arises when going to the data. The theory assumes we observe returns *ex ante*, when the decision to pursue the idea is made. But at the time the decision to pursue the idea is made, the revenue will generally be uncertain, and we only observe its *ex post* realization. Suppose that there are two outcomes, a favorable outcome yielding a return of R with probability p and an unfavorable one yielding no return. Then $\rho = pR$. As long as p does not depend on R, using R in place of ρ will overestimate revenues but will correctly compute elasticities.

Revenue from Authorship of Fiction Books. We now examine a particular category of creative individuals: authors of fiction books. Ideally one would like to observe revenues for various books for each author, to account for the possible *ex ante* uncertainty about *ex post* sales. Such data are not available, hence we proceed with what is. Although we do not have data on lifetime income of individual authors, we do have data on the revenue generated by individual book sales. We ignore the fact that it is costly to produce books once they are written, which is irrelevant to our ends insofar as the cost of producing each copy of a book is independent of the number of copies produced and sold. In summary, our assumptions are

- The opportunity cost of writing books is constant.
- Expected revenues from the sale of a "successful" book are perfectly anticipated, and the probability of failure does not depend upon the private return.
- The marginal cost of producing books is small relative to sales price.

Then income per unit of time taken to produce a book is $r = \lambda\phi\rho$ and, given current copyright laws, one can safely set $\phi = 1$ in what follows. We can compute the aggregate income of all authors who earn at least a given amount, $M^r(r)$, and of course $M(\rho) = (1/\lambda)M^r(\rho/\lambda)$ has the same elasticity. We gathered data on revenues for fiction books published in March and September of 2003 and 2004, respectively; our samples range between 1,200 and 1,300 books for each of these four months. The details of the data collection procedure can be found in Appendix 2. Figure 4.2 shows $M(\rho)$ computed on the basis of the September 2003 data. Figure 4.3 shows a plot on logarithmic axes, including a close-up to illustrate more clearly the increasing nature of the elasticity on both ordinary and logarithmic axes. The data for the other months, not reported but available, yield extremely similar results.

FIGURE 4.1. _

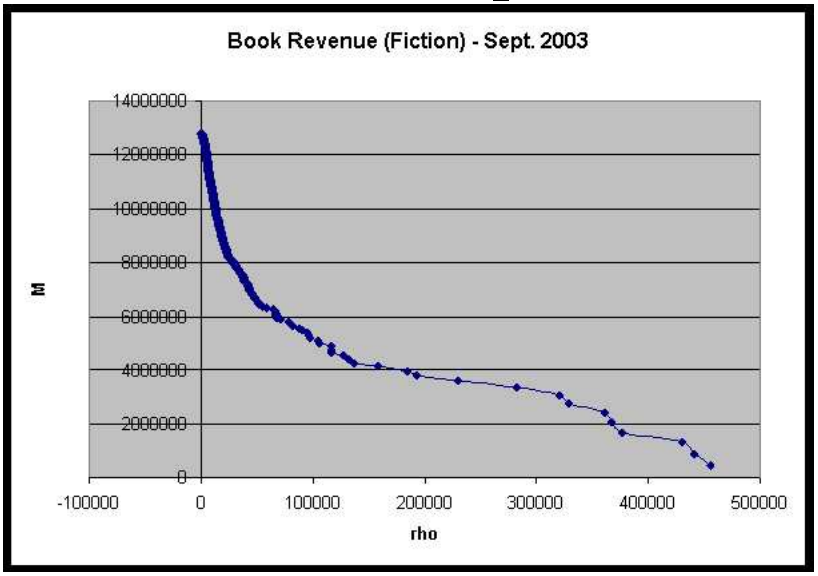
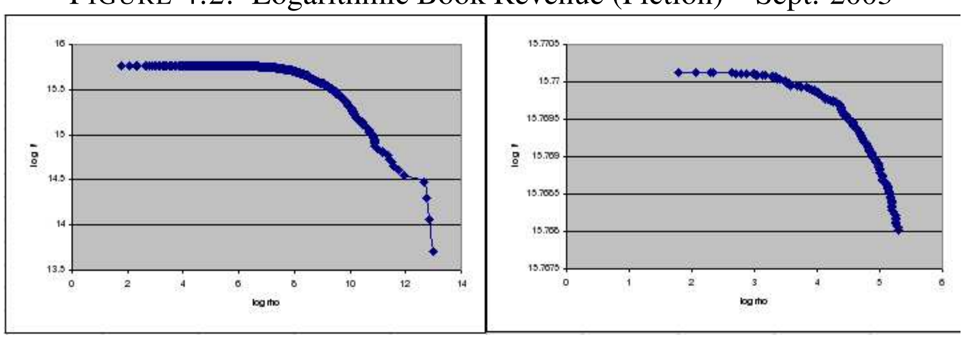


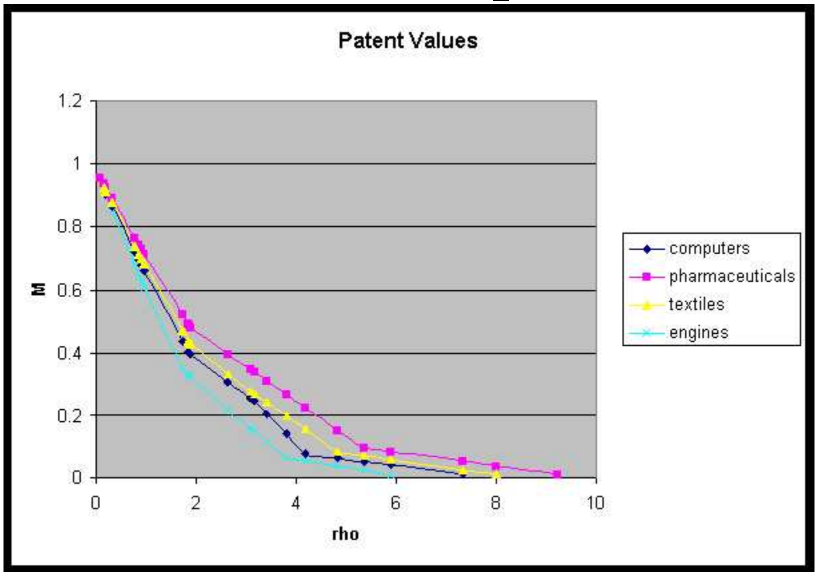
FIGURE 4.2. Logarithmic Book Revenue (Fiction) – Sept. 2003



Three comments are in order. First, for less successful books the $M(\rho)$ function is nearly linear, and overall the function exhibits increasing elasticity – a fact that can be seen more clearly in the logarithmic plots. The second striking feature is the discontinuity between roughly \$150,000 and \$300,000 in revenue.[4] This is broadly consistent with other data on books revenues: Leibowitz and Margolis [2003] report that less than 200 out of 25,000 titles account for roughly two-thirds of all book revenues. This is considerably more concentrated than we find in our data – but certainly reflects a strong discontinuity. These books appear to be predominately by

[4]The sales data are from a single distributor, Ingram, constituting about one-sixth of the book market, so total revenues would be about six times this number.

FIGURE 4.3. _



"big name" authors, who are largely irrelevant for optimal copyright policy: the relevant part of the $M(\rho)$ function is the part near the cutoff – that is, for marginal, not infra-marginal, books. The third fact is how small the fixed cost $wh(\rho)$ may be for writing and publishing fiction; in September 2003, 1,181 books, out of a total of 1,223, earned $50,000 or less – corresponding to total revenue of approximately $300,000. These books accounted for 50% of total revenue, that is, $6M out of $12M. The numbers for the other months are similar. In the same data, 984 books earned less than $10,000; hence our estimate of the marginal author's opportunity cost should be placed at $60,000 or less.

Patent Values. A similar analysis of the value of patents is possible – with the reservation that it is less likely for patents that *ex post* value can be anticipated *ex ante*. If we disaggregate by industry, it is at least plausible that the fixed cost of the innovation is not systematically related to the realized revenues. We use the value of patents estimated by Lanjouw [1993] for four German industries on the basis of patent renewal rates and data on the cost of renewal. We graph the corresponding $M(\rho)$curves in Figure 4.4. Were elasticity constant, the distribution would be Pareto. As can be seen, in no case are the tails similar to that of a Pareto distribution – the curves fall far too close to zero. Numerical estimates can be found in Figure 4.5. This reports for each industry and for increasing values of ρ, the elasticities evaluated at the midpoint of each segment of the linear spline. The number in square brackets is the corresponding value of $-\rho M_i'(\rho)$.

FIGURE 4.4. Elasticities

Computers	Pharmaceuticals	Textiles	Engines
.22 [.17]	.14 [.12]	.19 [.15]	.32 [.23]
.74 [.40]	.53 [.33]	.66 [.38]	.95 [.45]
.93 [.30]	.75 [.30]	.88 [.31]	1.12 [.32]
3.76 [.60]	2.35 [.48]	2.42 [.44]	3.04 [.42]
2.73 [.12]	2.81 [.16]	3.02 [.14]	3.37 [.12]

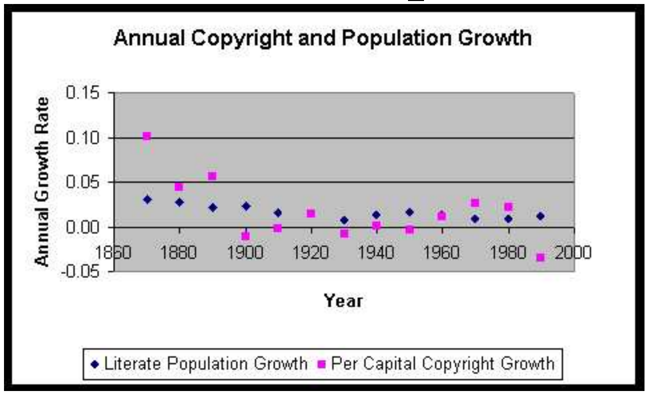
With the exception of the highest category of ρ for computers, elasticities are increasing everywhere. The values of $-\rho M_i'(\rho)$ are also relevant because the same ϕ applies across sectors. Hence the aggregate distribution is $M(\rho) = \sum_i M_i(\rho)$, where i indexes industries. Unfortunately, the fact that each $M_i(\rho)$ function has increasing elasticity does not imply that this is true for $M(\rho)$. However, if $M_i'(\rho)$ is increasing, then the corresponding elasticity is increasing as well, and increasing $-\rho M_i'(\rho)$ is a condition that does aggregate. While not always increasing, $-\rho M_i'(\rho)$ is increasing in the relevant range, that is, at lower values of ρ, for all i. This implies the elasticity of $M(\rho)$ is also increasing, at least for values of ρ near the threshold, which is what matters.

Our findings for patents appear to accord well with the existing empirical literature. To name but a few recent studies, Harhoff, Scherer, and Vopel (1997) use a data set of full-term patents applied for in 1977 and held by West German and U.S. residents. They compare the ability of various empirical distributions, including the Pareto, to fit the data and find that a two-parameter lognormal distribution provides the best fit; this implies increasing elasticity. Silverberg and Verspagen (2004) use a variety of different data sources from both Europe and the U.S.A. and two different measures of ρ (citations and monetary values). They find that, while the overall distributions are well approximated by exponential ones, it is the *upper tail* that is better captured by a Pareto distribution. Again, the exponential implies increasing elasticity. As our concern here is with the shape of the $\mu(\rho)$ near the lower cutoff value, this is supportive of our claim. The econometric literature on the value of patents, stemming from the paper of Pakes [1986] (see Hall, Jaffe, and Tratjenberg [2004] for a recent update and new results), seems to almost unanimously find that the appropriate distribution is a log-normal or an exponential, for both of which the elasticity of the total revenue function is increasing.

5. EMPIRICAL ANALYSIS OF LABOR DEMAND

As we have seen, there is a close connection between the elasticity of total monopoly revenue and labor demanded by the ideas sector. Here we

FIGURE 5.1. _



exploit this relationship to get a second source of information about whether the elasticity of total monopoly revenue is increasing or decreasing.

Copyright Time Series. First we apply our analysis of labor demand to a time series of U.S. copyright. Here we must assume that the distribution $M(\rho)$ is time invariant, and that ϕ is either constant or increasing over time – as in fact it is. We measure the scale of the market by the size of the literate population,[5] and the amount of labor in the sector by the number of copyright registrations. The relevant annual growth rates for the U.S. are reported, by decade, in Figure 5.1. If elasticity of total monopoly revenue is constant or decreasing, we expect to see per capita copyright growing more rapidly than population. This is in fact the case prior to 1900 and for 1970-80, but those are both anomalous periods. For the pre-1900 period one must notice that copyright registration only began in 1870, so the huge initial increase in registrations is unlikely to reflect a corresponding increase in the actual output of literary works. In particular, it is important to realize that in 1891 it became possible for foreign authors to get U.S. copyrights for the first time.[6] Similarly, in 1972 it became possible to copyright musical recordings other than phono records – previously such recordings were protected under other parts of the law. In 2000 6.8% of new copyrights were for sound recordings, so it is not surprising that copyright registrations jumped up 1972. In 1976, the term of copyright, which since 1909 had been 28 years, plus a renewal term of 28 years, was increased to the life of the author plus 50 years. In 1988 the United States eliminated the requirement of registering a copyright, so after that time, there is no reason to

[5]The literacy adjustment makes little difference; in 1870 when the copyright registration data begin, the literacy rate is already 80%, climbing to 92.3% by 1910.

[6]A brief history of U.S. copyright can be found at U.S. Copyright Office [2001a]. The 1972 change is described in U.S. Copyright Office [2001b].

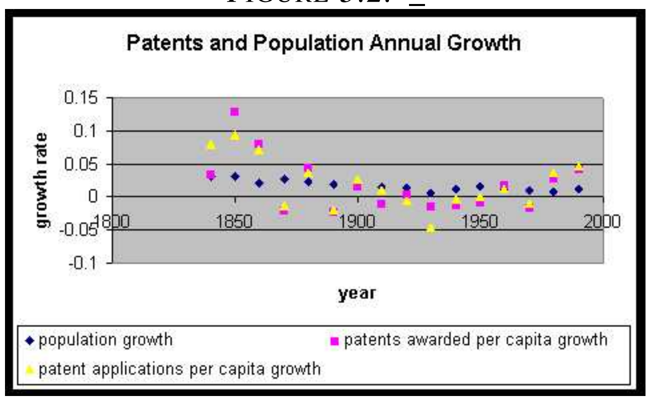
think of copyright registrations as a particularly good measure of the output of literary works.

What all this means is that we should focus on the period between the major copyright acts of 1909 and 1972. Here we find that overall the literate population grew by 92%, while the number of copyright registrations grew by only 12%. Moreover, the literate population grew faster than the per capita copyright registrations in every decade, although in 1920-1930 and 1960-1970 the two growth rates are very similar. This is especially dramatic because as we noted above, there was considerable technological change during the period, with entirely new areas such as movies, recorded music, radio, and television opening up: by 2000 only 48% of new copyright registrations were for literary works, while in 1909 literary works accounted for the bulk of copyright registrations. Further, while the number of copyright registrations in the U.S.A. *overestimates* the share of the U.S. per capita labor dedicated to literary work, the size of the literate population grossly *underestimates* the size of the relevant market. The first is because a large number of foreign writers register their work in the U.S.A., the second because the growth of per capita income and, especially, the expansion of "American culture" around the world greatly increased the potential market size.

Patent Time Series. We next turn to the demand for labor used to produce patentable ideas. One issue that arises is whether we should measure the scale of market λ by population or by GDP. Increases in per capita GDP increase the scale of the market, but they increase the opportunity cost of labor in the non-idea sector (working with existing ideas) by the same proportion, so have no impact on the effective scale of the market. On the other hand, increased productivity in the non-idea sector may also be reflected in increased productivity in the idea sector: double the per capita income may mean twice as many ideas. We will focus on population as a more conservative measure of λ in time series data, where per capita GDP is increasing. In the cross section we will examine both population and GDP as measures of scale of market.

Figure 5.2 is the patent analog of Figure 5.1 and is quite similar. Whether we measure patentable activity by patents awarded or by patent applications, from 1890 to 1980 the growth rate of per capita patents exceeds the growth rate of population in only two decades, 1900-1910 and 1960-1970, and in both cases by only a trivial amount. In other decades, the growth rate of patents per capita is much lower than population growth, in some cases even negative. Overall, from 1890 to 1980 population grew at a rate of 1.4% per year and per capita patents at 0.1% per year. Before 1890 patents per capita grew considerably faster than population, with a large drop in

FIGURE 5.2. _



patents from 1860 to 1870 most likely because the reform of the patent law and patent office in 1861 made it considerably more difficult to get a patent. In the opposite direction, in the period after 1980 it became much easier to get and enforce a patent – the landmark event in this period being the creation, in 1982, of a special federal court to try patent cases. In summary, the time series of patents lead us to the same conclusions we reached with copyright: that patents have grown less than market size, thereby suggesting that the elasticity of monopoly revenue is increasing also in this case.

An alternative to measuring either patent applications or awards is to use R&D expenditure as a proxy for the amount of labor used in creating new ideas. R&D expenditure, while in principle a better measure of input than patents, has a number of its own problems. First, the concept of R&D expenditure is fairly fuzzy and available only for relatively recent years – the major source of data being an NSF survey conducted since 1953. The definition used by the NSF is "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications." Firms and government agencies are surveyed and asked to report how much they spend on this activity.

The picture of R&D expenditure as measured by the NSF is ambiguous and yet different from that of the number of patents – ambiguous because the choice of *which measure* of R&D expenditure one should consider is not obvious. One possibility is to focus on the private sector only. However, we would expect that research financed by the federal government – much of which is carried out at private institutions – both produces useful ideas and increases the demand for skilled labor. On the other hand, there are reasons to believe that the federal expenditure in R&D reacts much less, or maybe not at all, to market incentives and to the expected profitability of innovations in particular.[7] Universities, either public or private, are obviously producing ideas and employing skilled workers, but the extent to which they respond to market incentives may have varied substantially during the last fifty years. In the light of this, we will report statistics for four aggregates: total, private sector plus universities, and these same two series adjusted for the wage rate of college and post-college workers. The latter are relevant because the wage skill premium increased dramatically during the last thirty years, and workers involved in R&D activities hold college, and most often post-college, degrees.

The ratio of total R&D expenditure to GDP has grown from 1.36% in 1953 to 2.78% in 2002, thereby doubling in fifty years. During the same time, population has grown about 80% and real GDP has almost quintupled. It may be worth noticing that the maximum value for the total R&D expenditure to GDP ratio, 2.88%, was reached in 1964. For the private plus universities aggregate, the same ratio has more than tripled between 1953 and 2002, going from 0.63% to 2.0%. Next, assume that the cost of labor employed in the idea sector grows, roughly, at one-half the college wage and one-half the post-college wage.[8] Then the cost of the average worker in the idea sector between 1963 and 2002, the period for which data are available, has grown by about 95%, while over the same period, the mean wage has grown by about 65%.[9] Between 1963 and 2002, the ratio of total R&D expenditure to GDP basically does not move, while the industry plus universities ratio goes from 0.9% to 2.0%. That is, the industry plus university ratio grows by 110%, population grows by 52%, and total GDP by 70%. Because our index of the relative wages in the idea sector has grown roughly 20% over the same period, it turns out that, if one uses total expenditure in R&D, then the share of workers in the idea sector has actually declined, implying a strongly increasing elasticity of $M(\rho)$; if, instead, one uses the private plus universities measure, it has grown by about 90%. The latter is somewhat higher than either the population or the GDP growth rates; hence, on the basis of the last index, one cannot rule out the

[7]This point is made by Jones [2004] while analyzing the R&D data and the "patent puzzle." However, we would expect some scale of market effect on federal R&D expenditure as well – as the scale of the market increases so does the tax base that pays for the expenditures.

[8]This is arbitrary but not unreasonable.

[9]High school graduate wages grew 20%, college graduate wages grew by 65%, and post-college graduate wages grew at 123%; see Eckstein and Nagypal [2004], Figures 1 and 3.

hypothesis that the elasticity of the total revenue function is either constant or decreasing.[10]

R&D Cross Section. Finally, we look at a cross section of countries. Here we run a simple cross-country regression with R&D as a fraction of GDP as the dependent variable and market size and the strength of IP protection as explanatory variables.[11] We initially assume that the domestic market is what is significant. If ℓ represents per capita labor effort in the ideas sector and we assume constant elasticity of labor demand with respect to market size, we can write $\log \ell = \vartheta \log(\phi\lambda)$. To account for the effect of both population and per capita GDP on market size, write $\lambda = y^{\alpha}N$, where N is population and y is per capita GDP. Ordinary least squares regression[12] gives $\vartheta = 0.20(0.03)$ and $\alpha\vartheta = .56(0.038)$, meaning that $\alpha = 2.8$, a remarkably large number that, if applied to the previous time series analysis would imply a strongly increasing elasticity.[13] The estimated elasticity with respect to λ is nowhere close to unity.

So far, we have assumed that the relevant market for R&D is the domestic market. More generally, we would measure $\lambda = \lambda_{domestic} + \lambda_{world}$, where λ_{world} is the fraction of world GDP available as a market for domestic R&D. Since regressing log R&D on λ gives essentially the same result as regressing on $\lambda_{domestic}/\lambda_{average}$, and regressing on $\log(\lambda_{domestic} + \lambda_{world})$ gives essentially the same result as regressing on $\lambda_{domestic}/\lambda_{world}$, the regression coefficient should be multiplied by $\lambda_{world}/\lambda_{average}$. Thus, if the ratio of revenue earned on R&D in foreign markets to domestic markets were on the order of 5, it would be possible for the elasticity of per capita R&D with respect to size of market to be near unitary. However, a ratio of 5 is implausibly large for most countries with the exception, possibly, of Switzerland and Luxembourg. Exports are almost everywhere a fraction, not a multiple of GDP. Consequently, a ratio of 5 would be possible only if R&D were much more intensive in export industries than the average – by a factor larger than 5. Using Lo's [2003] detailed data from Taiwan, in 1991 export intensive industries spent about 1.8 times as much on R&D

[10] An endless list of additional caveats should be added. The tax and accounting treatments of R&D have changed substantially over the period, favoring the relabeling of many sources of cost as R&D expenditure. The Cold War, and the changing federal policies toward basic research also add additional uncertainty to the interpretation of the data.

[11] To measure the latter we use an index developed by Walter Park, to whom we are grateful for providing us with his data. Details of the construction can be found in Park and Lippholdt [2003].

[12] Standard errors in parentheses, $R^2 = 0.65$.

[13] The underlying data include 34 countries for the period 1980-1997, and can be found at http://www.dklevine.com/data.htm.

as domestic-oriented industries. Using microdata on renewal rates to estimate the value of patents Lanjouw, Pakes. and Putnam [1998] find the highest value of the "implicit subsidy from patenting abroad" at 35% for the U.K. and Germany, with most countries receiving 15-20% of income from a patent from rights held abroad. So the evidence easily contradicts the idea that $\lambda_{world}/\lambda_{average}$ is on the order of 5.

6. Implications for IP

What consequences does our analysis have for the optimal IP policy? The first set of calculations indicates that IP protection for patents is probably too high, but this conclusion is somewhat tentative. In the case of copyright, it seems conclusive that copyright terms are far too long. The second set of calculations strongly indicates that the scale of market effect is quantitatively significant and that there should be substantial reductions in the length of IP term in response to size of market increases.

To turn this into operational policy prescriptions, the first step is to translate ϕ – our measure of effective IP protection – into the relevant policy parameter - the length of patent and copyright term. This depends on the interest rate and on depreciation.

Length of Term, Depreciation, and Effective Protection. Suppose that the real interest rate is r, that all ideas depreciate at a common rate d and that the length of term is T. Then – with perfect enforcement – the effective protection is $\phi = 1 - e^{-(r+d)T}$. Reasonable estimates of the real interest rate lie between 2% and 4%. Since the Sony Bono Copyright Term Extension Act of 1998, copyright protection in the U.S. is life of the author plus 70 years, or 90 years for works without an author. If we take the remaining life of an author to be roughly 35 years, this would mean 105 years of protection.[14] Current patent length in the U.S. for utility patents (inventions) is 20 years.

Depreciation rates are more difficult. In our data for books published in September 2003, revenues accrued during the four months of 2003 were 2.4 times those during the 10 months of 2004; meaning that per month sales fell by a factor of 6 over about one-third of a year, or an annual depreciation rate of nearly 95%.[15] Capital goods depreciation rates are generally thought to be close to 8% per year, including housing and building, which depreciate more slowly. Little data are available about the depreciation rate of ideas so,

[14]Akerloff et al. [2002] use an estimate of 30 additional years of life and a 7% real interest rate.

[15]This is consistent with data for the other months and with the general rule of thumb in the publishing industry that the most significant book sales occur within three months of publication.

FIGURE 6.1. Effective Copyright Protection

$r+d$	$T=20$	$T=105$
0.02	0.330	0.878
0.03	0.451	0.957
0.04	0.551	0.995
0.07	0.753	0.999
0.08	0.798	1.000
0.09	0.835	1.000
0.38	1.000	1.000

insofar as ideas correspond to capital vintages, they may well depreciate at the same rate; some very good ideas (the law of gravity) may not depreciate at all.

If the flow of sales is constant over time, for a copyright length of $T = 105$ years, and different interest rates r and depreciation rates d, the corresponding values of $\phi = 1 - e^{-(r+d)T}$ are given in Figure 6.1.

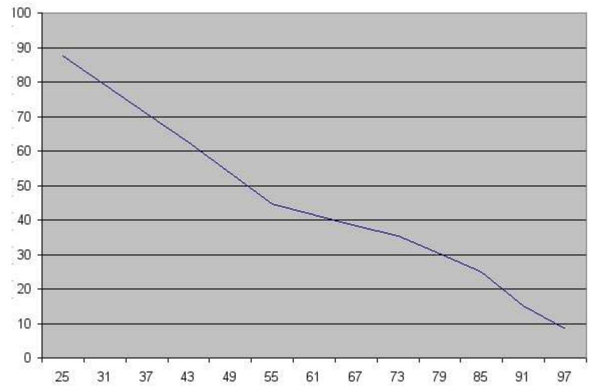
The low values 0.02, 0.03, 0.04 for $r+d$ correspond to no depreciation; the intermediate values 0.07, 0.08, 0.09 correspond to a modest depreciation rate of 5%; we do not report any values larger than 0.38 (that is, depreciation between 34% and 36%) since, even with just a 20-year term, $\phi = 1$ at this point. In summary, for realistic interest and depreciation rates, the current copyright term certainly corresponds to $\phi = 1$ in our model, while current patent terms correspond to roughly $\phi = 0.9$.

Calibration of Demand. To analyze the optimal level of protection, we need to know $\overline{\nu}^C$ and $\overline{\nu}^M$, besides Υ. A useful benchmark case is that of linear demand and constant marginal cost so that $\overline{\nu}^C = 2$ and $\overline{\nu}^M = 3/2$. But, is linear demand empirically relevant?

Take first the case of a small cost-saving innovation – for example, a way of making a machine work a little better. This is the type of thing most people think of when they think of an "invention," although only a small fraction of patents are of this type. Demand for a small cost-saving innovation is equal to the per machine cost saved up to the number of machines – then drops to zero. Since the innovation is small it has an insignificant effect on the number of machines. In this case, to a good approximation, $\overline{\nu}^C = \overline{\nu}^M = 1$, since we have normalized so that the monopoly profit is 1. In this case the elasticity of total monopoly revenue does not really matter: the social optimum is to set $\phi = 1$, and it does not change in response to the scale of market.

More generally, it is easy to see that if demand is concave, then $\overline{\nu}^M$ and $\overline{\nu}^C - \overline{\nu}^M$ are smaller than in the linear case – the extreme case being that

FIGURE 6.2. Demand Proportional to Income



U.S. Income Distribution 2001
Source: U.S. Census

of a small cost-saving innovation – while if demand is convex then $\overline{v}^M$, and $\overline{v}^C - \overline{v}^M$ are larger than in the linear case. Notice that larger $\overline{v}^M$ and $\overline{v}^C - \overline{v}^M$ increase the scale of market effect, but have an ambiguous effect on the level of IP: larger $\overline{v}^M$ tending to increase and larger $\overline{v}^C - \overline{v}^M$ tending to decrease optimal IP.

In understanding how good the benchmark linear case is, it is important to recognize that demand for most innovations is strongly affected by income. Take the case of new drugs: it is probably a good approximation to think of willingness to pay as proportional to individual income. From 2001 census data for the U.S., assuming that each individual demands one unit of an innovation, with willingness to pay proportional to income, we construct the demand curve shown in Figure 6.3. In other words, demand based on linear Engel's curves is, to a good approximation, linear. Artistic creations such as books, movies, and music are similar to drugs in that demand is heavily dependent on income. In fact drugs and artistic creations are undoubtedly superior goods, meaning that the fraction of income spent on them increases as income goes up.

If we start with linear demand and assume linear Engel's curves, then goods that are strongly superior, in the sense that the fraction of income spent on them rises at an increasing rate, have convex demand curves. Conversely for goods that are strongly inferior – orphan drugs are a likely example – demand will be concave.

The conclusion is that for most types of goods, the linear demand approximation is conservative – most likely overstating the level of optimal IP protection and understating the optimal rate of decrease in response to

FIGURE 6.3. Optimal Protection and Term Length

Υ	$\hat{\phi}$	$r+d=0.2$	$r+d=0.4$	$r+d=0.08$
0.03	0.13	7	4	2
0.10	0.24	14	7	4
0.15	0.33	20	10	5
0.20	0.40	26	13	7
0.30	0.51	36	18	9
0.40	0.60	46	23	12

market size. The exception is in the case of small cost-saving innovations – which to a certain extent matches the idea of "process" rather than "product" patents in patent law. Historically, "process" patents – patents on methods for doing things – have received stronger protection than "product" patents. The theory indicates that this is in fact the right approach. Unfortunately, despite the great historical success – for example, in the development of the chemical industry – of allowing only "process" patents in countries such as Germany, the Anglo-French system of allowing products the same protection as processes has become widespread.

The Static Optimum. To determine the optimal level of protection we can solve the first order condition from the Appendix to find

$$\hat{\phi} = \left(\frac{1}{\overline{\nu}^C} + \frac{\overline{\nu}^C - \overline{\nu}^M}{\overline{\nu}^C} \frac{(1+\Upsilon)}{\Upsilon} \right)^{-1}.$$

For elasticities, marginal books in the book revenue data have an elasticity of about 0.03; marginal inventions in the patent data have elasticities ranging as high as 0.32. In Figure 6.2 we report (second column) the optimal values of $\hat{\phi}$ corresponding to elasticities in the empirically range from 0.03 to 0.40. The other two columns translate the optimal $\hat{\phi}$ in lengths of term, using different interest and depreciation rates.

Two facts stand out. First, optimal length of protection is less than 1 – meaning that, given that elasticity is increasing, optimal copyright and patent protection should decline with the size of the market. Second, in the case of copyright, optimal copyright length is much less than actual copyright length; since the actual cutoff value of ρ in the data is quite small, even an elasticity of 0.05 may be a tremendous overestimate of the actual elasticity on the margin. Certainly it is hard to justify as few as 7 years of copyright based on this data; if we consider depreciation – not in the empirical range of 95%, but say in the range of 5% – copyright protection should be at most several years. This is generally consistent with our scale of market calculations below under the hypothesis that 28 years at the start of the 20th century was about right.

In the case of patents, estimated elasticities appeared somewhat larger, with .20 being a sensible middle ground estimate. With a real interest rate plus depreciation rate of 4%, this implies an optimal patent length of 13 years, while with a more realistic depreciation adjustment it would be closer to 7 years - again, not so terribly different than what we would get if we assumed term length were correct at the beginning of the twentieth century and imputed the increase in market size. If we took the high end elasticity of .4 and a real interest rate of just 2%, the optimal term would be 46 years; hence it is not impossible, at least in principle, to reconcile existing patent term with available data. Realistic estimates, though, suggest that optimal patent term should be between 7 and 13 years.

The Scale of Market Effect. To examine the scale of market effect, we differentiate the first order condition from the Appendix to find

$$\mathfrak{E}\phi(\lambda) = -\frac{1}{1+(1+\Upsilon)/\mathfrak{E}\Upsilon+(1/(\overline{\nu}^C-\overline{\nu}^M))}.$$

To get a feeling for this, note that in the simple and empirically relevant case that $M(\rho)$ is linear $\mathfrak{E}\Upsilon = 1+\Upsilon$. Consequently $\mathfrak{E}\phi(\lambda)$ is $-1/2$ or less negative depending on $\overline{\nu}^C-\overline{\nu}^M$. When demand is linear, $\overline{\nu}^C-\overline{\nu}^M = 1/2$, and $\mathfrak{E}\phi(\lambda) = -1/4$. This means that a 10% increase in size of market should reduce effective protection by 2.5%. For example, if the world economy is growing at 4% per year, then a simple rule of thumb would be to reduce protection by about 1% per year. In the case of 20-year patents that would mean about two months each year. One implication of this is that during the last century in which world GDP grew by a factor of roughly 40, optimal protection should have declined from 20 years to about 1 year.

A paradigmatic case is that of popular music. Forty years ago, at the time of Elvis Presley and the Beatles, new recordings selling a million units were considered exceptional successes and awarded "golden records," while in the current times a successful record sells easily ten or twenty million copies. The effective size of the market has, therefore, increased at least a factor of ten. At the same time, advances in recording and digital technologies have reduced the fixed cost required to produce a new record to about one-fifth of its earlier level. This suggests that the socially optimal length of copyright protection should have dropped by about a factor of twelve. Unfortunately, in the case of copyright, terms have been moving in the opposite direction; copyright terms have grown by a factor of about four since early in the twentieth century. This means that, at least for recorded music, they currently are on the order of a hundred times longer than they should be. A similar calculation can be performed for books and movies. Consider the fact that, since the beginning of the past century, world GDP has grown by nearly two orders of magnitude. It is reasonable to argue that the size of the market for books and movies must have grown at least as much as literacy has surged, and the availability of playing devices has increased more than proportionally due to the dramatic drop in their relative prices. Hence, if the copyright term of 28 years at the beginning of the 20th century was socially optimal, the current term should be a little over a year, rather than the current term of approximately 100 years. This gives a ratio of 100 between the actual copyright terms and their socially optimal value.

7. Conclusion

For the first time, to the best of our knowledge, we merge established theory of IP protection with available data on the value of innovations to quantify the socially optimal term of IP protection, and its relation with market size. Among existing models, we use the one in which IP has the potentially highest social value; hence our calculations are likely to overestimate the optimal length of IP protection. We draw information from a wide variety of independent empirical sources, both time-series and cross section, to calibrate the model, and always reach the same set of policy conclusions.

- The elasticity of total monopoly revenue is increasing, hence the term of IP protection should decrease over time as the market size increases. Our best estimate, given the historical growth rate of market size, is that IP protection terms should decrease of about two months per year.
- Current copyright and patent terms are equivalent to complete monopoly protection for the full economic life of new goods, and are dramatically higher than optimal ones, sometime by two orders of magnitude.
- On the basis of the available evidence, our best estimate of the length of optimal copyright term is about one year, and that of patents is about seven to thirteen years.

References

[1] Akerloff, G., K. Arrow, T. Bresnahan, J. Buchanan, R. Coase, L. Cohen, M. Friedman, J. Green, R. Hahn, T. Hazlett, C. Hemphill, R. Litan, R. Noll, R. Schmalensee, S. Shavell, H. Varian, and R. Zeckhauser (2002), *Amici Curiae in Support of Petitioners in the Supreme Court of the United States, Eldred versus Ashcroft*.

[2] Boldrin, M. and D. K. Levine (1999), "Perfectly Competitive Innovation," University of Minnesota and UCLA, November.

[3] Boldrin, M. and D.K. Levine (2002), "The Case Against Intellectual Property," *The American Economic Review (Papers and Proceedings)* **92**, 209–212.

[4] Boldrin, M. and D.K. Levine (2003), "Rent Seeking and Innovation," *Journal of Monetary Economics*, **51**, 29–41.

[5] Boldrin, M. and D.K. Levine (2004), "Lawrence Klein Lecture: The Case Against Intellectual Monopoly," *The International Economic Review,* **45,** 327–350.

[6] Boldrin, M. and D.K. Levine (2005), "The Economics of Ideas and Intellectual Property," *Proceedings of the National Academy of Sciences*, **102**, 1252–1256.

[7] Eckstein, Z. and E. Nagypal (2004), "U.S. Earnings and Employment Dynamics 1961-2002: Facts and Interpretations," mimeo, University of Minnesota and Northwestern University, January.

[8] Grossman, G. M. and E. Helpman (1991), "Trade, Knowledge Spillovers and Growth," *European Economic Review (Papers and Proceedings)*, **35,** 517–526.

[9] Grossman, G. M. and E. Helpman (1994), "Endogenous Innovation in the Theory of Growth," *Journal of Economic Perspectives*, **8,** 23–44.

[10] Grossman, G. M. and E. Helpman (1995), "Technology and Trade," in G. Grossman and K. Rogoff, eds., *Handbook of International Economics*, vol. III, North Holland.

[11] Grossman, G. M. and E. L. Lai (2004), "International Protection of Intellectual Property," *American Economic Review* **94**, 1635–1653.

[12] Hall, B. (2001), "NBER Patents Data File," http://elsa.berkeley.edu/users/bhhall/pat/datadesc.html.

[13] Hall, B., A. Jaffe and M. Tratjenberg (2004), "Market Value and Patents Citations," *Rand Journal of Economics*, forthcoming.

[14] Harhoff, D., F.M. Scherer, and K. Vopel (1997), "Exploring the Tail of Patented Invention Value Distributions," discussion paper FS IV 97-27, WZB, Berlin.

[15] Jones, C. (2004), "Growth and Ideas," prepared for the *Handbook of Economic Growth*.

[16] Lanjouw, J. (1993), "Patent Protection: Of What Value and How Long?" NBER Working Paper 4475.

[17] Lanjouw, J., A. Pakes, and J. Putnam (1998), "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data," *Journal of Industrial Economics,* **46**: 405-432.

[18] Leibowitz, S. and S. Margolis (2003), "Seventeen Famous Economists Weigh in on Copyright: The Role of Theory, Empirics, and Network Effects," University of Texas at Dallas.

[19] Lo, S. (2003), "Strengthening Intellectual Property Rights: Experience from the 1986 Taiwanese Patent Reforms," UCLA.

[20] Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, **54**, 755–784.

[21] Park, W. and D. Lippholdt (2003), "The Impact of Trade-Related Intellectual Property Rights on Trade and Foreign Direct Investment in Developing Countries," OECS, http://www.american.edu/cas/econ/faculty/park/TD-TC-WP-2003-42-final.pdf.

[22] Romer, P. M. (1990), "Endogenous Technological Change," *Journal of Political Economy*, **98**, s71-s102.

[23] Silverberg, G. and B. Verspagen (2004), "The Size Distribution of Innovations Revisited: An Application of Extreme Value Statistics to Citation and Value Measures of Patent Significance," MERIT working paper 2004-021, Maastricht University.

[24] U.S. Copyright Office (2001a), *A Brief History and Overview*, Circular 1a, http://www.copyright.gov/circs/circ1a.html.

[25] U.S. Copyright Office (2001b), *Copyright Registration for Sound Recordings*, Circular 56, http://www.copyright.gov/circs/circ56.html.

APPENDIX 1: PROOFS

Proposition. 3.1. *Suppose that the opportunity cost of skilled labor is constant at* $w = 1$*. When* $\hat{\phi}(\lambda) < 1$*, in a neighborhood of* $\rho = 1/\lambda\hat{\phi}(\lambda)$*, the following three cases hold. (I)* $\Upsilon'(\rho) > 0$ *implies* $\hat{\phi}(\lambda)$ *is unique and strictly decreasing; (II)* $\Upsilon'(\rho) = 0$ *implies* $\hat{\phi}(\lambda)$ *is unique and constant; and (III)* $\Upsilon'(\rho) < 0$ *and* $\hat{\phi}(\lambda)$ *unique*[16] *implies* $\hat{\phi}(\lambda)$ *is strictly increasing.*

Proof. Divide welfare W by $\lambda g(\lambda)$, then differentiate it with respect to ϕ to get the first order condition for a social optimum

$$\begin{aligned} FOC(\lambda,\phi) &= \\ &= -\left[(1/\phi)\left\{\phi\overline{\nu}^M + (1-\phi)\overline{\nu}^C\right\} - 1\right](1/\lambda\phi)M'(1/\phi\lambda) \\ &\quad - (\overline{\nu}^C - \overline{\nu}^M)M(1/\phi\lambda). \end{aligned}$$

Divide through by $M(1/\phi\lambda) > 0$, the resulting expression

$$NOC(\lambda,\phi) = \left[(1/\phi)\left\{\phi\overline{\nu}^M + (1-\phi)\overline{\nu}^C\right\} - 1\right]\Upsilon(1/\lambda\phi) - (\overline{\nu}^C - \overline{\nu}^M)$$

has the same qualitative properties as $FOC(\lambda,\phi)$: it has the same zeros, the same sign on the boundary, and $NOC_\phi(\lambda,\phi) < 0$ is sufficient for a zero to be a local maximum.

We next differentiate with respect to ϕ to find the second order condition for a social optimum

$$\begin{aligned} NOC_\phi &= \\ &- \left[(1/\phi)\left\{\phi\overline{\nu}^M + (1-\phi)\overline{\nu}^C\right\} - 1\right](1/\lambda\phi^2)\Upsilon'(1/\lambda\phi) \\ &- \frac{\overline{\nu}^C}{\phi^2}\Upsilon(1/\lambda\phi). \end{aligned}$$

The second term is unambiguously negative. The first term has two factors of interest. We have $(1/\phi)\left\{\phi\overline{\nu}^M + (1-\phi)\overline{\nu}^C\right\} - 1$ representing social surplus of the marginal idea produced; since privately it yields zero profit, it must yield positive social surplus. If the other factor $\Upsilon'(1/\lambda\phi) > 0$ then there is a unique solution to the social optimization problem; if

[16] In this case we cannot guarantee that the second order condition is satisfied, so we must rule out the possibility that $\hat{\phi}(\lambda)$ has multiple values.

$NOC(\lambda,1) \geq 0$, then that solution is $\hat{\phi}(\lambda) = 1$; otherwise it is the unique solution to the first order condition $NOC(\lambda,\phi) = 0$.

In the latter case, we may use the implicit function theorem to compute

$$\frac{d\phi}{d\lambda} = -\frac{NOC_\lambda}{NOC_\phi} \propto NOC_\lambda =$$
$$= -\left[(1/\phi)\left\{\phi\overline{v}^M + (1-\phi)\overline{v}^C\right\} - 1\right](1/\lambda^2\phi)\Upsilon'(1/\lambda\phi),$$

which has the opposite sign to $\Upsilon'(1/\lambda\phi)$. □

Proposition. 3.2. *Consider two different aggregate monopoly revenue functions M_1, M_2 that have the same value $M_1(\rho) = M_2(\rho)$ and derivative $DM_1(\rho) = DM_2(\rho)$ (hence, elasticity $\Upsilon_1(\rho) = \Upsilon_2(\rho)$) at ρ. If $D\Upsilon_1(\rho') < D\Upsilon_2(\rho')$ for $\rho' \geq \rho$, then*

(1) *Labor demand associated to M_1 is smaller than the one associated to M_2; that is,*
$$\int_\rho^\infty -[DM_1(\rho')/\rho']d\rho' < \int_\rho^\infty -[DM_2(\rho')/\rho']d\rho'.$$
(2) *The elasticity of labor demand associated to M_1 is greater than the elasticity of labor demand from M_2; that is, $\mathfrak{E}[H_1(\rho)] > \mathfrak{E}[H_2(\rho)]$.*
(3) *As the elasticity of total revenue goes from increasing, to constant, to decreasing, the elasticity of the associated labor demand functions increases monotonically.*

Proof. **Step 1:** $M_1(\rho') > M_2(\rho')$

Here and in what follows, $\rho' \geq \rho$ holds. Then, $D\Upsilon_1(\rho) - D\Upsilon_2(\rho) < 0$ by assumption. Moreover

$$D\Upsilon(\rho) = D[-\rho DM(\rho)/M(\rho)] =$$
$$= \frac{1}{\rho}[\Upsilon(\rho) + \Upsilon^2(\rho)] - \rho D^2M(\rho)/M(\rho)$$

so $D^2M_2(\rho) - D^2M_1(\rho) = (M(\rho)/\rho)[D\Upsilon_1(\rho) - D\Upsilon_2(\rho)] < 0$, where $M(\rho)$ is the common value of M_1 and M_2 at ρ. Then, for ρ' near ρ we have

$$M_1(\rho') - M_2(\rho') \approx (1/2)[D^2M_1(\rho) - D^2M_2(\rho)](\rho' - \rho)^2 > 0$$

Moreover, if $M_1(\rho'') - M_2(\rho'') < 0$ for some larger ρ'', then $M_1(\rho') - M_2(\rho') = 0$ for some $\rho'' > \rho' > \rho$, since both functions are continuous. Let $\hat{\rho}'$ be the smallest such ρ', that is, the first point to the right of ρ where M_1 and M_2 cross. Then $\Upsilon(\hat{\rho}') = -\rho' DM(\hat{\rho}')/M(\hat{\rho}')$ and the assumption that $\Upsilon_1(\hat{\rho}') < \Upsilon_2(\hat{\rho}')$ imply $DM_1(\hat{\rho}') > DM_2(\hat{\rho}')$, that is, M_1 crosses M_2 from below, which is impossible since to the left of $\hat{\rho}'$ we already know that $M_1 > M_2$.

Step 2: $\int_\rho^\infty -[DM_1(\rho')/\rho']d\rho' < \int_\rho^\infty -[DM_2(\rho')/\rho']d\rho'$

Recall that $M(\infty) = 0$. Integration by parts gives

$$\int_{\rho}^{\infty} -[DM(\rho')/\rho']d\rho' = -M(\rho')/\rho'|_{\rho}^{\infty} - \int_{\rho}^{\infty} M(\rho')/(\rho')^2 d\rho' =$$

$$= M(\rho)/\rho - \int_{\rho}^{\infty} M(\rho')/(\rho')^2 d\rho'$$

from which

$$\int_{\rho}^{\infty} -[DM_1(\rho')/\rho']d\rho' - \int_{\rho}^{\infty} -[DM_2(\rho')/\rho']d\rho' =$$

$$= -\int_{\rho}^{\infty} [M_1(\rho') - M_2(\rho')]/(\rho')^2 d\rho' < 0$$

Step 3: $\mathfrak{E}[H_1(\rho)] > \mathfrak{E}[H_2(\rho)]$
Because

$$\mathfrak{E}[H(\rho)] = \mathfrak{E}[\int_{\rho}^{\infty} -[DM(\rho')/\rho']d\rho'] =$$

$$= \frac{-\rho DM(\rho)/\rho}{\int_{\rho}^{\infty} -[DM(\rho')/\rho']d\rho'} =$$

$$= \frac{-DM(\rho)}{\int_{\rho}^{\infty} -[DM(\rho')/\rho']d\rho'}.$$

$\mathfrak{E}[H_1(\rho)]$ and $\mathfrak{E}[H_2(\rho)]$ have the same numerator, and, because of Step 2, the first has a smaller denominator. Hence the conclusion. □

APPENDIX 2: DATA

Book Revenue. We collected all the titles, ISBN numbers, and sale prices listed by www.amazon.com for the query *hardcover fiction books* and for the four publication periods of March and September 2003 and 2004. The sales data are from the Ingram stock statistics, automatic telephone line at 615-213-6803. The Ingram stock statistics system gives the following statistics for each ISBN number punched in: "Total sales this year," "Total sales last year," "Total current unadjusted demand," "Total last week demand." Total revenue for each book is calculated using the total sales data from Ingram and the November 2004 sales price listed on www.amazon.com. Ingram is a large book distributor, and generally thought to generate roughly one-sixth of all book sales. It should be noted that the sales prices on www.amazon.com are changing over time, most often decreasing, so we might have underestimated the revenue during the first year for books published during September 2003. Because of the large number of observations, we do not reproduce the data here, but it is available from http://www.dklevine.com/data.htm.

Copyright Time Series. The basic source of the copyright registration time series is from the annual report of the copyright office from 2000, which can be found at http://www.copyright.gov/reports/annual/2000/appendices.pdf. This also includes the breakdown of registrations by type for 2000. Population data for 1901-1999 is from the U.S. Census http://www.census.gov/population/estimates/nation/popclockest.txt; data prior to 1901 is from http://www.census.gov/population/censusdata/table-2.pdf; the two sources have a slight discrepancy for the 1900 population with the former source reporting 76,094,000 and the latter (which we used) 76,212,168. The year 2000 data was from the 2000 Census. Literacy rates are from http://www.arthurhu.com/index/literacy.htm. The data we used can be found at http://www.dklevine.com/data.htm.

Patent Time Series. R&D Expenditures by Sectors: National Patterns of R&D Resources: 2002 Data Update, Table D, National Science Foundation GDP: National Income and Production Account, Table 1.1.5, Bureau of Economic Analysis. Population: 1953-1959: Population Estimates Program, Population Division, U.S. Census Bureau, release date: April 2000 1960-2002: U.S. Census Bureau, Statistical Abstract of U.S., 2004-2005.