

MARKET SIZE AND INTELLECTUAL PROPERTY PROTECTION

MICHELE BOLDRIN AND DAVID K. LEVINE

Washington U. in Saint Louis, Fed. Res. Bank of Saint Louis, and CEPR; Washington U. in Saint Louis, and NBER

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 relative 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 SEJ2005-08783-C04-01 (Ministry of Education, Spain). Our ideas benefited from comments received during presentations at Columbia University, Dallas, Richmond and San Francisco Federal Reserve Banks, Chinese University of Hong Kong, ASU, Rochester, 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. We are also grateful to two anonymous referees for their detailed comments.

Corresponding Author: David K. Levine, Department of Economics, Washington University in Saint Louis, Saint Louis, MO 63130; Ph:1-314-935 5648. Email: david@dklevine.com.

First Version: 2nd December, 2003, This Version: March 3, 2008.

1. INTRODUCTION

In this paper we study the main factors 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 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 the impact of changes in market size, finally, we put everything together to obtain the implications for optimal policy.

Conventional wisdom¹ is that optimal policy involves a trade-off between increasing the monopolistic distortion on inframarginal ideas, and increasing the number of usable ideas by innovating at the margin. To understand this trade-off requires a general equilibrium model. The best current model of this type is that of Grossman and Lai [2004], which builds on earlier work by Grossman and Helpman [1991, 1994, 1995] studying innovation in a Dixit-Stiglitz framework. To carry out measurement in the Grossman-Lai framework, we introduce a more detailed model of the production function for ideas. In our setup, we assume that 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. For a given level of protection, ideas will be produced if they exceed a cutoff level of private return. We can think of this cutoff as the marginal idea.

The key policy issue is how many ideas there are of different qualities: if lowering the cutoff brings many marginal ideas into the market it will be worth the increased monopoly distortion on infra-marginal ideas, while if it brings few marginal ideas into the market it will not. In other words, policy depends on how ideas of different qualities are distributed. In particular, the aggregate revenue earned by innovators with private returns that are above the cutoff serves, in our, analysis as a kind of cumulative distribution function for the quality of ideas. The key policy parameter is the elasticity of this revenue function with respect to the quality of the marginal idea. In particular, increasing revenue elasticity with respect to the marginal idea

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

implies that optimal protection decreases with the scale of the market and conversely.

Our primary empirical goal is to measure the elasticity of revenue with respect to the marginal idea. We use two methods of measurement. First, we attempt to measure the elasticity directly using data we collected on the distribution of book revenues and existing estimates of the value of patents. Second, the elasticity of revenue with respect to the marginal idea has implications for the demand for skilled labor. When this elasticity is constant or decreasing, an increase in market scale will cause the demand for labor to increase even more rapidly. 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.

Each of our empirical efforts has many caveats and each individual numerical calculation must be regarded with a healthy degree of skepticism. However, despite the fact that each uses very different data, all of our analyses lead to the same conclusion. The heart of our empirical analysis is contained in four graphs. Two of them report a measure of revenues plotted against an index of private returns, both in logarithmic units. Increasing revenue elasticity with respect to the marginal idea means that this curve should be concave. It is, and it is extremely concave: caveats aside, the data would have to be remarkably defective if the true underlying curves failed to be concave. The graphs in the second set plot the growth rate of ideas per capita against the growth rate of population. Constant or decreasing revenue elasticity with respect to the marginal idea has the implausible implication that the growth rate of per capita ideas should be at least as large as – in practice much larger than – the growth rate of the scale of the economy. Here the evidence is strong that per capita ideas do not grow faster than population. We conclude that there is robust evidence that revenue elasticity with respect to the marginal idea is increasing and consequently that protection should optimally decrease with the scale of the market. As the scale of the market has been growing over time due to increased population, income, and trade, this calls into doubt policies of recent decades, such as those of the WIPO, which have led us in the opposite direction.

2. THE MODEL

Grossman and Lai [2004] provide a fully articulated dynamic equilibrium model of the trade-off between increasing the monopolistic distortion on inframarginal ideas, and increasing the number of usable ideas by innovating at the margin. A key result of that paper is that the equilibrium is a steady state, and a steady state that may be readily analyzed by static methods. For this reason we do not reproduce their entire model here, but rather focus on

the static analysis and disaggregate their technology for the production of ideas from human capital and labor to make it amenable to measurement.²

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.³ An idea is produced if, given the patent length ϕ , the prospective monopolist finds it profitable to pay the cost of innovation.

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 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

²Details of how to embed the model here and theirs can be found in the longer working paper Boldrin and Levine [2005b].

³We 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.

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 simply $H(\rho) = \int_{\rho}^{\infty} h(\rho')d\rho'$, which is similar to a cumulative distribution function, except in reverse. Notice however, that as $\rho \rightarrow 0$ it may be that $H(\rho) \rightarrow \infty$, that is, we do not assume it is necessarily feasible to produce all possible ideas.

The function $H(\rho)$ is then a measure of total labor input, while aggregate 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 revenue with respect to the marginal idea*⁴ as $\Upsilon(\rho) \equiv -\rho DM(\rho)/M(\rho) > 0$. 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 consider more closely the private and social benefit from an idea. A new idea makes it possible to produce new goods and services. Suppose that Q is the quantity of goods and services produced using a given idea, and that these goods and services are produced at constant marginal cost c . For example, the idea may be a new novel, and Q the number of copies that are produced; or the idea may be the design of an electronic component, and Q are the number of components produced. Suppose individual (per capita) demand at price p is given by $q(p)$. Notice the assumption that individual demand does not depend on the scale of the economy.⁵ Then $Q = \lambda q(p)$, and we may invert this function to obtain the marginal willingness to pay function $p = P(Q/\lambda)$.

⁴In Boldrin and Levine [2005b] we show that this is identical to the elasticity of research output with respect to labor in the Grossman and Lai [2004] formulation.

⁵This might not be the case if there were network effects in using the output.

The monopolist maximizes $[P(Q/\lambda) - c]Q$. Denote the solution p^M for the monopoly price and Q^M for the monopoly quantity. The private benefit of the innovation to the monopolist is $(p^M - c)Q^M$. The social value of the innovation depends on whether or not there is competition. Under competition output is determined by $P(Q^C/\lambda) = c$. Notice that Q^M, Q^C are independent of the scale of the market. If $Q^i \in \{Q^M, Q^C\}$ social value is $\int_0^{Q^i} (p(Q'/\lambda) - c)dQ' = \lambda \int_0^{Q^i/\lambda} (p(q) - c)dq$. Notice that the ratio of social to private benefits $(Q^M/\lambda)(p^M - c)/\int_0^{Q^i/\lambda} (p(q) - c)dq$ depends only on per capital consumption of the idea under competition and monopoly. These are independent of the scale of market λ .

It is instructive to consider the case of linear demand, so that $p(q) = a - bq$. Normalizing the cost of the invention to one, the return to monopoly is $\lambda\rho = \lambda(a - c)^2/2b$, the social value of the idea in the absence of monopoly is 2ρ and in the presence of monopoly it is $3\rho/2$. The key fact is that with linear demand the ratio of social to private return is independent of the private return. We will take this as our basic hypothesis. That is, under monopoly, we assume that the social return is $\bar{v}^M\rho$, while once the monopoly expires, the social return is $\bar{v}^C\rho$, where $\bar{v}^C > \bar{v}^M > 1$.⁶ With this assumption of proportionality, 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:

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

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) $DY(\rho) > 0$ implies $\hat{\phi}(\lambda)$ is unique and strictly decreasing; (II) $DY(\rho) = 0$ implies $\hat{\phi}(\lambda)$ is unique and constant; and (III) $DY(\rho) < 0$ and $\hat{\phi}(\lambda)$ unique⁷ implies $\hat{\phi}(\lambda)$ is strictly increasing.*

⁶To the extent that the ratio of social to private return increases with private return our results about market scale are reinforced: there is little reason to encourage ideas of marginal private value as their social value is so low. To the extent that the ratio of social to private return decreases with private return, intellectual property is not terribly appealing as a system for encouraging new ideas: in the extreme case where the correlation is negative, it results in the ideas with lowest social value being produced at the expense of those with the highest social value.

⁷In 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.

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

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

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, in the constant elasticity case as the size of the economy grows, the *share* of workers in the idea sector grows as well. The next proposition, proven in Appendix 1, shows how this finding can be extended from the case of constant elasticity to decreasing elasticity of the total monopoly revenue.

Proposition 3.2. *Consider two 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 $\mathbb{E}[H_1(\rho)] > \mathbb{E}[H_2(\rho)]$.*
(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 demand 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 demand 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 $\mathbb{E}[g(\lambda)]$ 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 technical restrictions discussed earlier for $h(\rho)$. 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. We are interested in the *ex ante* distribution of the expected return on an idea. In the data, we observe *ex post* a proxy for revenue such as patent value or book revenue, and we do not observe either the opportunity cost of each inventor or the labor cost of his ideas. These are important limitations. We address them in two ways: first we look for data in which there is little *ex ante* uncertainty about returns, and second we look for data in which the cost of producing ideas is similar for different ideas. In examining patent data, there is little we can do about the problem of *ex ante* versus *ex post* returns, but we attempt to restrict attention to ideas with

comparable production cost by examining data from narrow sectors which produce goods that are relatively homogeneous. On the copyright front, we examine data from trade fiction for two reasons: first, the opportunity cost of producing a novel is reasonably homogeneous across individuals, and second, after the first novel, the earnings from subsequent novels is relatively predictable based on past sales.

We first examine the data assuming that we can measure ρ without error and that the cost of each invention is the same. Because this is certainly not true, at the end of the section we examine the consequences for measurement of both return uncertainty and cost that is correlated with return. We show that the consequent biases are not likely to be severe.

A second issue concerns aggregation. The same ϕ applies across many industries. Aggregating across industries, we see that $M(\rho) = \sum_i M_i(\rho)$, 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 $-\rho DM_i(\rho)$ is increasing, then the corresponding elasticity is increasing as well, and this condition does aggregate. For this reason we examine not only whether the elasticities are increasing, but also whether $-\rho DM_i(\rho)$ is increasing.

Patent Values. There is a substantial literature analyzing the value of a patent - what we refer to as the private return. If we aggregate across industries, some industries such as pharmaceuticals have patents that have much higher private returns and much higher private costs. By examining estimates that are disaggregated by industry, we hope that the fixed cost of the innovation is not so systematically related to the realized revenues. The most suitable data we have been able to identify is the value of patent data from Lanjouw [1993] for four German industries. These estimates are constructed on the basis of patent renewal rates and data on the cost of patent renewal.

From Lanjouw's data, we graph the corresponding $M(\rho)$ curves in logarithms in Figure 1. Notice that the tails are not at all similar to that of a Pareto distribution – the curves fall far too close to zero. Numerical estimates can be found in Figure 2. 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 $-\rho DM_i(\rho)$.

With the exception of the highest category of ρ for computers, elasticities are increasing everywhere. While not always increasing, $-\rho DM_i(\rho)$ is increasing in the relevant range, that is, at lower values of ρ , for all i .

Our findings for patents appear to accord well with the existing empirical literature. To name but a few recent studies, Harhoff, Scherer, and

FIGURE 1.

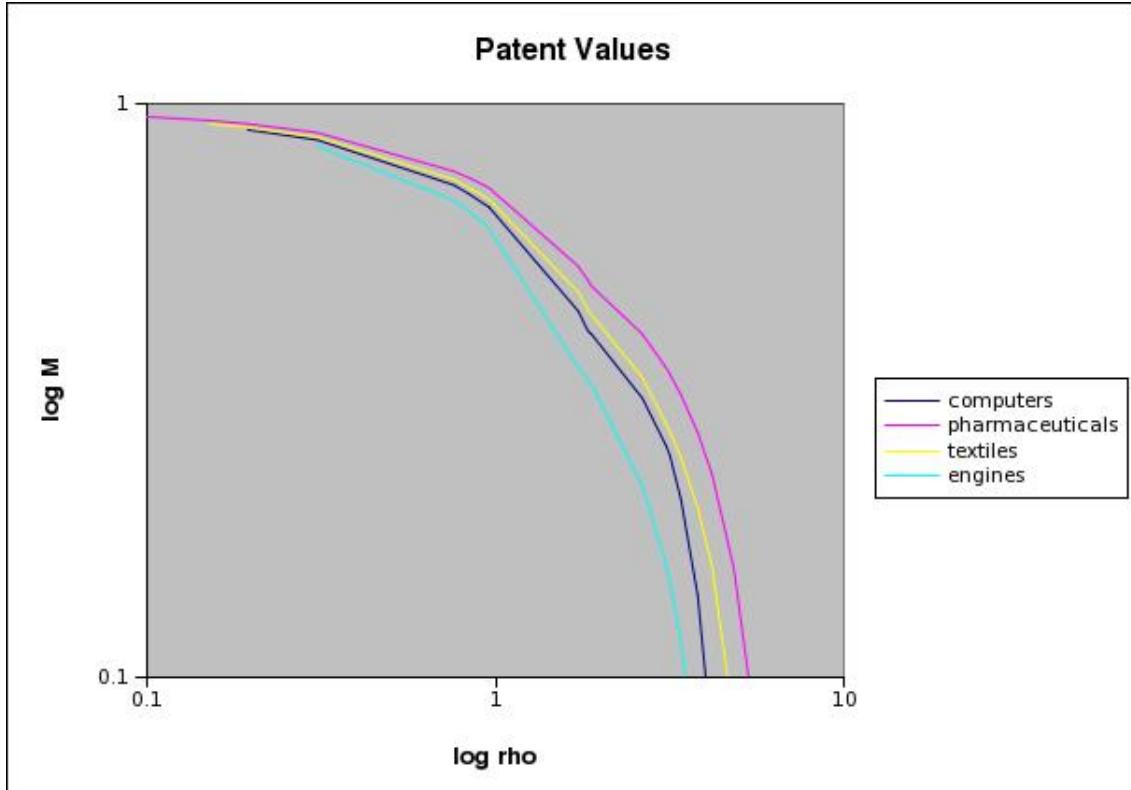
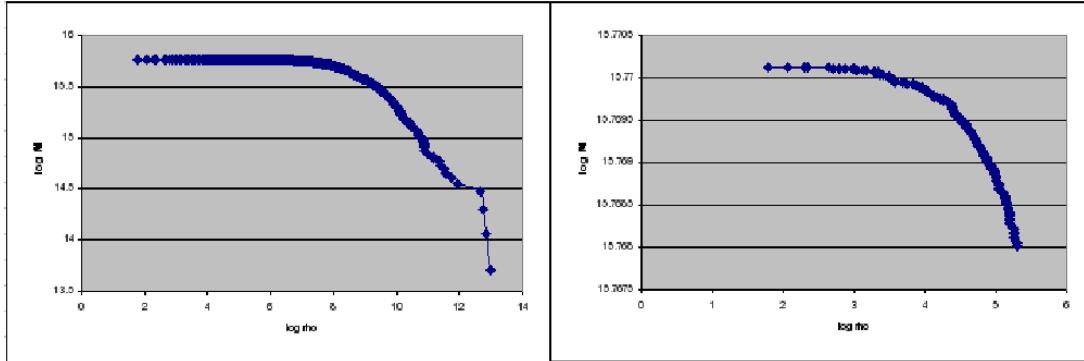


FIGURE 2. Elasticities

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]

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. 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. As our concern here is with the shape of the $\mu(\rho)$ near the

FIGURE 3.



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.

Revenue from Authorship of Fiction Books. We now examine a particular category of creative individuals: authors of fiction books. Although we do not have data on lifetime income of individual authors, we do have data on the revenue generated by individual book sales. While it is costly to print and distribute books once they are written, authors are traditionally paid a percentage of gross revenue, so that revenue is an accurate measure of the private return to the author.

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. The sales data are from a single distributor, Ingram, constituting roughly one-sixth of the book market, so total revenues are about six times the numbers reported. Figure 3 shows $M(\rho)$ computed on the basis of the September 2003 data in logarithmic units, including a close up of the lower revenue section of the plot. The data for the other months, not reported but available, yield extremely similar results.

There are two key issues in analyzing this data. First, as is evident in graph, the data have a discontinuity between roughly \$150,000 and \$300,000

in revenue.⁸ Needless to say, the \$150,000 and \$300,000 books are predominately by “big name” authors, who are largely irrelevant for optimal copyright policy: the relevant part of the $M(\rho)$ function is for marginal, not inframarginal, books. Second, many books earn surprisingly little revenue. In September 2003, 1,181 books, out of a total of 1,223, earned \$50,000 or less through Ingram – 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 through Ingram – corresponding to total revenue of \$60,000 or less. It should not be thought that such a book is a “commercial failure” – the fixed cost in printing were already quite low by 2003, and publishers are often happy to make a profit by selling, for example, a small run of a thousand or less monographs on economic theory to libraries. Nor is it necessarily a failure for an author who may earn 15% of the \$60,000, or around \$9,000 for what may be as little as a month of effort. However, as we get closer to the lower end of the data, it is clear that many books are published that are not commercial successes. Examining specific titles convinces us that these are largely vanity press items. This does raise the question of whether a model of profit maximizing authors is the correct model with which to understand fiction books. It is true that if the author pays to have the book published, he receives no benefit from copyright. However, if the publisher pays the author, even if the author receives some additional psychological benefit from seeing his work in print, the model applies. Hence, while works below the cutoff of commercial viability are published the model applies to those above the cutoff.

In analyzing the specific details of the elasticity we face the problem of knowing where exactly the cutoff for commercial viability is. However, the results are not terribly sensitive to this detail. Except near the irrelevant upper tail, the function clearly exhibits increasing elasticity. The actual elasticity Υ ranges from zero when $\rho = 0$ to $\Upsilon = 0.1$ when $\rho = 5000$ – which is a plausible value for books that are not mere vanity press items – to about $\Upsilon = 0.3$ for the higher values of ρ , so the value of Υ is quite low regardless of where the cutoff is.

Robustness and Measurement Error. As we noted above, there are problems with measurement of both *ex ante* private return and with the assumption that the opportunity cost of invention is independent of the return. To

⁸This 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. That is more concentrated than we find in our data but also reflects a strong discontinuity.

analyze the matter, let us distinguish between r the observed proxy return, and ρ the actual *ex ante* expected return.

Let us take up the issue of *ex ante* uncertainty first. Particularly in the patent data, much of the observed private return may not have been known *ex ante* when the cost of developing the idea was incurred. What consequence does this have for our estimates? First consider the simple case where the only uncertainty is the possibility of failure. Then the *ex post* return can take on two values, 0 with probability π_F and r with probability $1 - \pi_F$. In other words, $\rho = (1 - \pi_F)r$ and this has no consequence for either the level or derivative of the elasticity. This model is especially relevant for pharmaceutical, where the market for the new product is relatively predictable, but there is a high probability it may be a failure in clinical trials.

Second, suppose that the realized return is $r = \theta\rho$, where θ is a positive random variable independent of ρ and having mean equal to one. In this case the observed revenue function is a convolution of the underlying revenue function we are trying to measure. This raises an issue similar to that of aggregating across different industries. In particular, elasticities do not aggregate, so we cannot conclude, for example, that if the elasticity is declining with the marginal idea, then this will remain true when we take the convolution. However, as we observed, $-\rho DM_i(\rho)$ does aggregate. So if elasticity declines with the marginal idea, then $-\rho DM_i(\rho)$ must decline even more rapidly, and therefore for the observed revenue function $M^r(r)$ the corresponding derivative $-rDM^r(r)$ must also decline. As we indicated, in the patent data, this is not the case.

Next, consider the consequences of opportunity cost that increases with observed return. This means that observed returns will be overstated for high returns, and understated for low returns. A simple model is

$$\rho = a + br.$$

In this case, $\rho' > \rho$ if and only if $r' > \frac{\rho-a}{b}$. It follows that

$$M(\rho) = M^r\left(\frac{\rho-a}{b}\right).$$

$$-\frac{\rho DM}{M} = -\frac{1}{b} \frac{\rho DM^r}{M^r} = -\frac{\rho}{\rho-a} \frac{rDM^r}{M^r}.$$

This implies that the actual elasticity is bigger than the observed elasticity and that the actual elasticity increases less than the observed elasticity. In particular, to the extent that costs are considerably higher for high value patents, our results are biased towards intellectual protection terms that are too short and decline too rapidly with the size of the market. It is worth noting, however, that the effect is not generally quantitatively great: in particular, if the observed revenue is linear, then the actual revenue will also

exhibit elasticity that increases with respect to the marginal idea. Moreover, in the important case of pharmaceuticals, a crucial component of the cost is the clinical trials - the cost of which does not depend on the value of the idea.

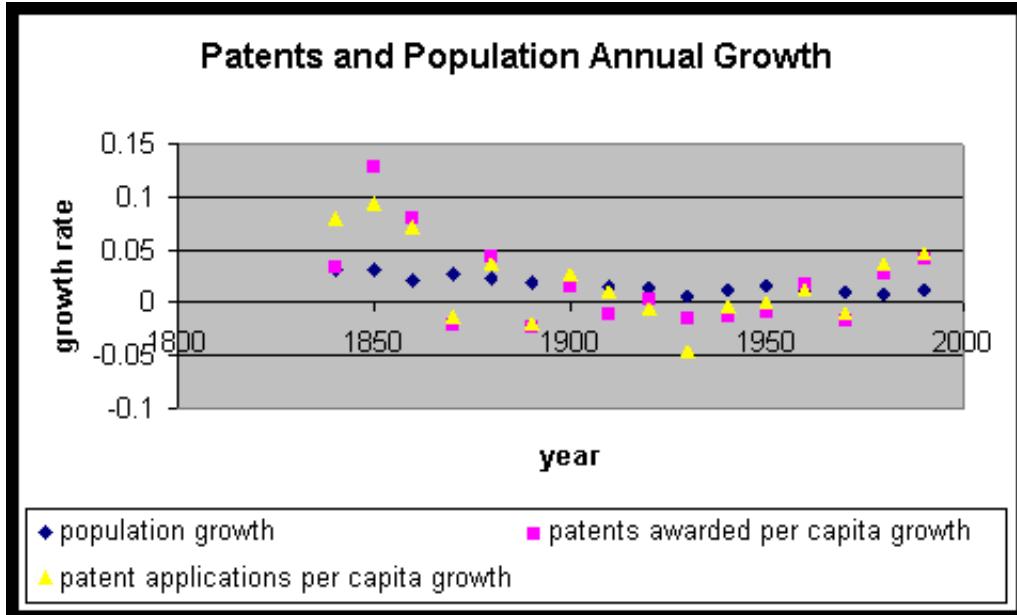
5. EMPIRICAL ANALYSIS OF LABOR DEMAND

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

Patent Time Series. We 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 increases in GDP have an ambiguous 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 per capita, or half the labor time required before to implement a new idea. For these reasons 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 4 reports annual growth of population against the per capita growth in patents awarded and patent applications. If elasticity of revenue with respect to the marginal idea is constant or decreasing, we expect to see per capita invention growing more rapidly than population. In fact, for per capita invention to grow at only the rate of growth of the scale of the economy, it must be that $E[g(\lambda)] = 0$, meaning that the number of available ideas does not increase at all as the size of the economy expands. This is not especially plausible, so if the elasticity of revenue with respect to the marginal idea is constant or decreasing, we would expect the growth rate of inventions to be much greater than the growth rate of the scale of the economy. As can be seen in Figure 4 this is far from the case. 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

FIGURE 4. —



per capita grew considerably faster than population, with a large drop in 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 continually easier to get and enforce a patent: during these decades software and business practices became patentable for the first time, and the standard of “obviousness” of a patent declined markedly. Some landmark events in this period being the creation, in 1982, of a special federal court to try patent cases and the State Street Decision in 1998 establishing the patentability of business practices. We should note also that this surge in patenting did not result in a surge in the growth of TFP as it should if these patents represent an economically meaningful increase in inventive activity. 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.⁹ 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.¹⁰ 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%.¹¹ Between 1963 and 2002, the ratio of

⁹This 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.

¹⁰This is arbitrary but not unreasonable.

¹¹High 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.

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 hypothesis that the elasticity of the total revenue function is either constant or decreasing.¹² On the other hand, while we cannot rule out this case completely, the combination of assumptions – including the assumption that the number of ideas do not grow as more people are added to the economy – needed to get this result are extremely implausible.

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.¹³ We initially assume that the domestic market is what is significant.

Let ℓ represents per capita labor effort in the ideas sector, recall that λ is the size of the market and the fraction of the market on which profits can be earned is ϕ . So from the private sector point of view, $\phi\lambda$ is the relevant size of the market. If we assume constant elasticity of labor demand with respect to relevant market size, we can write $\log \ell = A + \vartheta \log(\phi\lambda)$, where ϑ is the elasticity and A is an uninteresting intercept term. Naturally both population N and per capita GDP y have an impact on market size, so we take $\lambda = y^\alpha N$. Our OLS equation is then $\log \ell = A + \vartheta \log \phi N + \alpha \vartheta \log y$.

Our sample consists of 35 countries every five years from 1980 to 1995.¹⁴ Data on per capital GDP (y), population (N), and R&D per GNI (ℓ) is from the World Bank. The Park index of patent protection (ϕ) ranges in value

¹²An 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.

¹³To 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 Lippoldt [2003].

¹⁴Some data is missing so there are 122 observations. The data and a detailed explanation is available from <http://www.dklevine.com/data.htm>.

from 0.33 in Indonesia prior to 1995 to a high of 4.86 in the U.S. in 1995.¹⁵ The result of simple OLS regression (standard errors in parentheses, $R^2 = 0.65$) is $\vartheta = 0.19(0.03)$ and $\alpha\vartheta = .56(0.038)$. Implicitly $\alpha = 2.8$.

The estimated elasticity of R&D with respect to λ is clearly not close either economically or statistically to unity. The estimate $\alpha = 2.8$ is a remarkably large number. It implies that as per capita income increased, market size expanded much more rapidly. Since, as in our earlier time series, R&D per capita increased very little during a period in which per capita income skyrocketed we are forced to the low value of $\vartheta = 0.19$.

One question that arises is whether this is an artifact of the assumption that the relevant market for R&D is the domestic market. First, we observe that the evidence suggests that the domestic market is the more important measure of market scale. Except for a few small rich countries such as Switzerland and Luxembourg¹⁶ exports are almost everywhere a fraction of GDP. It is true that export intensive industries are generally more R&D intensive than domestic industries - in Taiwan, using Lo's [2003] detailed data, in 1991 export intensive industries spent about 1.8 times as much on R&D as domestic-oriented industries. This does still not suggest a large effect of the sales of ideas abroad on market size. 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 the income from a patent from rights held abroad.

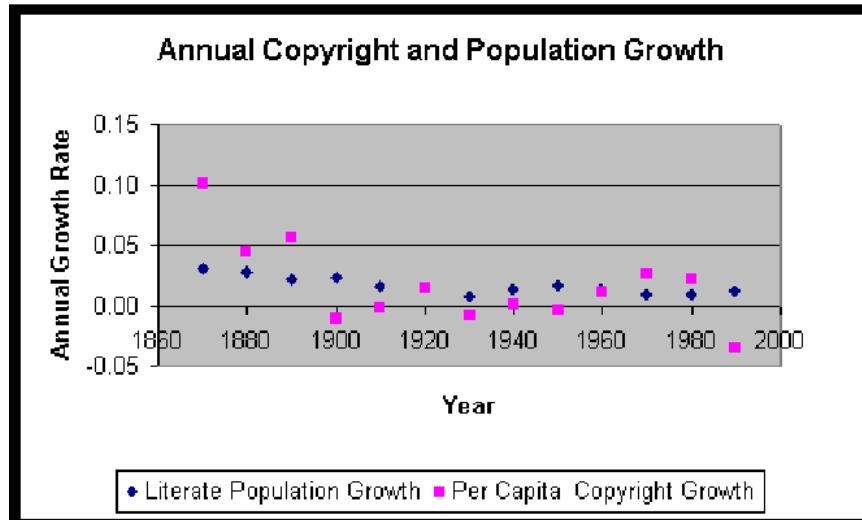
Another angle of attack on the issue of the role of the world market is to observe that the scale of the world market is growing at essentially the same rate for all countries in any given year, so the omitted variable of world demand should be highly correlated with a time trend. However adding a time trend has virtually no effect on the estimation. The coefficient estimates and standard errors are virtually unchanged $\vartheta = 0.19(0.03)$ and $\alpha\vartheta = .55(0.041)$ while the annual effect of growth in world GDP on R&D per capita is estimated to be only 0.3%. This time trend is not only economically insignificant, but it has a t-statistic of only 0.37, so the data is consistent with the absence of a time trend as well.

Copyright Time Series. Finally, 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

¹⁵Note that since we are taking logs, the exact units of ϕ do not matter.

¹⁶Switzerland is in our data, Luxembourg is not.

FIGURE 5. —



literate population,¹⁷ 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. 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 begins 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.¹⁸ 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 in 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 think of copyright registrations as a particularly good measure of the output of literary works.

¹⁷The 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.

¹⁸A 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].

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.

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, instead, 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 i , 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 $\phi = 1 - e^{-(i+d)T}$. 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

FIGURE 6. Effective Copyright Protection

$i + d$	$T = 20$	$T = 105$
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

protection.¹⁹ 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%.²⁰ 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, 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 $\phi = 1 - e^{-(i+d)T}$ are given in Figure 6.

The low values 0.02, 0.03, 0.04 for $i + d$ 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, as well as Υ , we need to know \bar{V}^C and \bar{V}^M . A useful benchmark is the base case of linear demand, in which case we showed that $\bar{V}^C = 2$ and $\bar{V}^M = 3/2$. But, is linear demand empirically relevant?

¹⁹Akerloff et al. [2002] use an estimate of 30 additional years of life and a 7% real interest rate.

²⁰This is consistent with data for the other months and with the general claim that the most significant book sales occur within three months of publication.

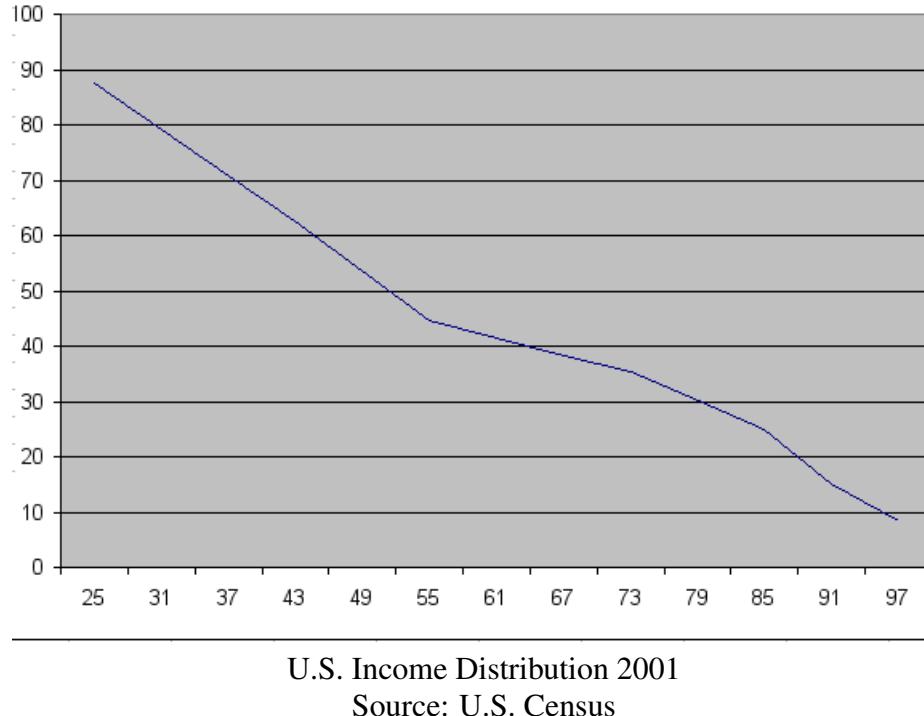
Take first the case of a small cost-saving innovation – for example, a way of making a machine work a little better. This is the type of thing most people think of when they think of an “invention,” although only a small fraction of patents are of this type. Demand for a small cost-saving innovation 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. On the other hand, this is also the case in which the innovation is usually “embodied” in some tool or machine-part, which is costly to be produced and cannot be easily imitated; said differently: this is the case the standard model is most incapable of capturing, and in which the Boldrin and Levine [1999] analysis applies best. Assuming away these concerns, we then conclude that, to a good approximation, $\bar{v}^C = \bar{v}^M = 1$ in this case, since we have normalized so that the monopoly profit is 1. When this applies, the elasticity of total monopoly revenue does not really matter as there is no deadweight loss in moving from competition to monopoly: 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 \bar{v}^M and $\bar{v}^C - \bar{v}^M$ are smaller than in the linear case – the extreme case being that of a small cost-saving innovation – while if demand is convex then \bar{v}^M , and $\bar{v}^C - \bar{v}^M$ are larger than in the linear case. Notice that larger \bar{v}^M and $\bar{v}^C - \bar{v}^M$ increase the scale of market effect, but have an ambiguous effect on the level of IP: larger \bar{v}^M tending to increase and larger $\bar{v}^C - \bar{v}^M$ 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 7. 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.

FIGURE 7. Demand Proportional to Income



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 market size. The exception is in the case of small cost-saving innovations that are completely disembodied and imitable – 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 now widespread.

A second issue we should highlight is that we have followed Grossman and Lai in specifying a static model of demand. In practice changing the length of protection may introduce intertemporal substitution – for example very short copyright periods might induce people to simply wait to buy the book for cheaper. It is important to bear in mind that what is relevant is the

FIGURE 8. Optimal Protection and Term Length

Υ	$\hat{\phi}$	$i+d = 0.2$	$i+d = 0.4$	$i+d = 0.08$
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

flow of services. In the case of patented inventions – pharmaceuticals especially – there is not likely to be much intertemporal substitutability between the flow of services from an invention. With creative goods such as books, there clearly is a great deal of intertemporal substitutability. However, there is also evidence that for these types of goods, consumers are very impatient. For example, the bulk of a books sales occur within a year of the time the book comes out; people pay a substantial premium for express delivery; the paperback comes out a year after the hardback, and so forth. The implication of this is that despite intertemporal substitutability, even very short copyright terms may represent a very high level of effective protection.

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

$$\phi = \left(\frac{1}{\bar{v}^C} + \frac{\bar{v}^C - \bar{v}^M}{\bar{v}^C} \frac{(1+\Upsilon)}{\Upsilon} \right)^{-1}.$$

In Figure 8 we report (second column) the optimal values of $\hat{\phi}$ corresponding to elasticities Υ in the range 0.03 to 0.40. Notice that the relevant elasticities correspond to the cutoff value of ρ which, strictly speaking, should be the lowest value seen in the data. In the copyright data the elasticity falls as low as zero, but the lower values of ρ clearly correspond to vanity press items. The lowest cutoff that is plausible for a “profit maximizing” work is around $\rho = 2000$ which gives an elasticity of 0.03. In the patent data, the cutoff values of ρ correspond to the first row of Figure 2. The highest value there is 0.32 corresponding to Engines. So we view the range 0.03 to 0.40 as the empirically relevant range. The third to fifth columns translate the optimal $\hat{\phi}$ in lengths of term, using different interest and depreciation rates.

Two facts stand out. First, optimal level of protection $\hat{\phi}$ is less than 1 – meaning that, given that elasticity is increasing, optimal copyright and patent protection should strictly decline with the size of the market. Second,

in the case of copyright, the optimal length is much less than the actual 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 .15 being the smallest sensible estimate. With a real interest rate plus depreciation rate of 4%, this implies an optimal patent length of 10 years, while with a more realistic depreciation adjustment it would be closer to 5 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 5 and 10 years.

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

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

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 $\bar{v}^C - \bar{v}^M$. When demand is linear, $\bar{v}^C - \bar{v}^M = 1/2$, 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. 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 because 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, since the path-breaking work of Nordhaus [1969], 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. On the basis of the available evidence, our best estimate of the length of optimal copyright term is about two years, that of patents is about ten years and IP protection terms should decrease at about two months per year.

Among existing models, we use the one in which IP has the potentially highest social value. This is likely to lead to an over-estimate the optimal length of IP protection. On the other hand, limitations in measurement may lead our calculations to under-estimate the length of IP protection. As we draw information from a wide variety of independent empirical sources, both time-series and cross section, our core conclusions are quite robust:

- The elasticity of total monopoly revenue is increasing, hence the term of IP protection should decrease over time as the market size increases. Our plots of revenue against private returns in logarithmic units show that this curve is sufficiently concave that the data would have to be remarkably defective for the true underlying curves to fail to be concave. Our plots of the growth rate of ideas per capita against the growth rate of population show that per capita ideas do

not grow faster than population, and certainly do not grow much faster than population, as the opposite claim would have it.

- The current copyright term is equivalent to complete monopoly protection and is dramatically higher than the optimal one.

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," forthcoming, *Journal of Monetary Economics*.
- [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 (2005a), "The Economics of Ideas and Intellectual Property," *Proceedings of the National Academy of Sciences*, **102**, 1252–1256.
- [7] Boldrin, M. and D.K. Levine (2005b), "IP and Market Size," available online at <http://www.dklevine.com/papers/scale22.pdf>.
- [8] 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.
- [9] Grossman, G. M. and E. Helpman (1991), "Trade, Knowledge Spillovers and Growth," *European Economic Review (Papers and Proceedings)*, **35**, 517–526.
- [10] Grossman, G. M. and E. Helpman (1994), "Endogenous Innovation in the Theory of Growth," *Journal of Economic Perspectives*, **8**, 23–44.
- [11] 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.
- [12] Grossman, G. M. and E. L. Lai (2004), "International Protection of Intellectual Property," *American Economic Review* **94**, 1635–1653.
- [13] Hall, B. (2001), "NBER Patents Data File," <http://elsa.berkeley.edu/users/bhhall/pat/datadesc.html>.
- [14] Hall, B., A. Jaffe and M. Tratjenberg (2004), "Market Value and Patents Citations," *Rand Journal of Economics*, forthcoming.
- [15] 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.
- [16] Jones, C. (2004), "Growth and Ideas," prepared for the *Handbook of Economic Growth*.
- [17] Lanjouw, J. (1993), "Patent Protection: Of What Value and How Long?" NBER Working Paper 4475.
- [18] 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.

- [19] 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.
- [20] Lo, S. (2003), "Strengthening Intellectual Property Rights: Experience from the 1986 Taiwanese Patent Reforms," UCLA.
- [21] Nordhaus, W.D. (1969), *Invention, Growth and Welfare: A Theoretical Treatment of Technological Change* Cambridge, Mass.: MIT Press.
- [22] Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, **54**, 755–784.
- [23] Park, W. and D. Lippoldt (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>.
- [24] Romer, P. M. (1990), "Endogenous Technological Change," *Journal of Political Economy*, **98**, s71-s102.
- [25] 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.
- [26] U.S. Copyright Office (2001a), *A Brief History and Overview*, Circular 1a, <http://www.copyright.gov/circs/circ1a.html>.
- [27] 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) $D\Upsilon(\rho) > 0$ implies $\hat{\phi}(\lambda)$ is unique and strictly decreasing; (II) $D\Upsilon(\rho) = 0$ implies $\hat{\phi}(\lambda)$ is unique and constant; and (III) $D\Upsilon(\rho) < 0$ and $\hat{\phi}(\lambda)$ unique²¹ 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

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

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

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

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.

²¹In 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.

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

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

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

$$\begin{aligned} \frac{d\phi}{d\lambda} &= -\frac{NOC_\lambda}{NOC_\phi} \propto NOC_\lambda = \\ &= - \left[(1/\phi) \left\{ \phi \bar{v}^M + (1-\phi) \bar{v}^C \right\} - 1 \right] (1/\lambda^2 \phi) D\Upsilon(1/\lambda \phi), \end{aligned}$$

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

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, $\mathbb{E}[H_1(\rho)] > \mathbb{E}[H_2(\rho)]$.
- (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

$$D\Upsilon(\rho) = D[-\rho DM(\rho)/M(\rho)] =$$

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

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

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

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

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

from which

$$\begin{aligned} \int_{\rho}^{\infty} -[DM_1(\rho')/\rho'] d\rho' - \int_{\rho}^{\infty} -[DM_2(\rho')/\rho'] d\rho' &= \\ &= - \int_{\rho}^{\infty} [M_1(\rho') - M_2(\rho')]/(\rho')^2 d\rho' < 0 \end{aligned}$$

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

Because

$$\begin{aligned} \mathbb{E}[H(\rho)] &= \mathbb{E}\left[\int_{\rho}^{\infty} -[DM(\rho')/\rho'] d\rho'\right] = \\ &= \frac{-\rho DM(\rho)/\rho}{\int_{\rho}^{\infty} -[DM(\rho')/\rho'] d\rho'} = \\ &= \frac{-DM(\rho)}{\int_{\rho}^{\infty} -[DM(\rho')/\rho'] d\rho'}. \end{aligned}$$

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

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.