INTELLECTUAL PROPERTY AND THE SCALE OF THE MARKET

MICHELE BOLDRIN AND DAVID K. LEVINE

Department of Economics University of Minnesota (Boldrin); UCLA and the Federal Reserve Bank of Minneapolis (Levine)

ABSTRACT. Intellectual property protection involves a tradeoff between the undesirability of monopoly and the desirable encouragement of creation and innovation. As the scale of the market increases, due either to economic growth, or the expansion of intellectual property rights through treaties such as the World Trade Organization, this tradeoff changes. We show that generally speaking, the socially optimal amount of protection decreases as the scale of the market increases.

Keywords: Economic Theory, Intellectual Property, Industrial Organization, Monopoly.

JEL Classification: X15; X16

We thank the National Science Foundation for financial support. Boldrin also acknowledges research support from the Spanish BEC2002-04294-C02-01. Our ideas benefited from comments at the Theory Workshop at Columbia University, the brown bag workshop at the Dallas Federal Reserve Bank, the Theory Workshop at ASU, and from a discussion with Kyle Bagwell. Thanks are also due to Hengjie Ai research assistance in collecting the book revenues 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: 30th November 2004.

1. INTRODUCTION

The U.S. Constitution gives Congress the power to "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." This recognizes the two basic economic features of intellectual property protection: on the one hand exclusive rights create monopoly power and so should be limited in time. On the other, monopoly power provides an incentive for creation and innovation. For practical reasons the same time limit applies across a wide variety of different creations and innovations: In U.S. law, copyright is life of author plus 70 years for individual works; 95 years for works for hire; design patents are 20 years and ornamentation patents are 14 years. Since the private profitability of creating and innovating varies widely, this means that for any fixed time limit many ideas will earn profits above and beyond the level needed to recoup the cost of innovation. In a larger market profits will be greater, and inframarginal ideas will earn additional economically unnecessary rents. Hence, as the market expands, it becomes possible to reduce the length of term without reducing the production of new ideas. But, as the market expands, some ideas that were not profitable to produce will become so, and reducing the length of term will discourage these marginal entrepreneurs. Which of these two competing forces should matter more for good policy?

In this paper we look at the general equilibrium interaction that determines how optimal protection varies with the scale of the market. Profitability of an innovation depends upon three factors: the initial cost of discovery, the elasticity of demand and, finally, the size of the market; all these elements vary widely and unsystematically across innovations. We concentrate on market size for three reasons. Contrary to the other two factors, market size is straightforward to measure; secondly, growth in per capita income and the expansion of international trade have increased market size by two orders of magnitude since when our patent and copyright legislation was first introduced. Finally, the ongoing process of trade expansion through has put the international harmonization of intellectual property rights at center stage, through the WTO-TRIPS agreement.

Our basic result is an intuitive one. Optimal copyright and patent term length involves a tradeoff between increasing the distortionary impact of unnecessary monopoly on inframarginal ideas, and increasing the number of marginal ideas. As the scale of the market increases, it will generally be desirable to give up some of the additional marginal ideas in exchange for reduction of monopoly across the broad variety of inframarginal ideas that will be produced anyway, and so the optimal policy should reduce the length of protection as the scale of the market increases.

We make this point in the context of a simple model, intentionally designed to render intellectual property protection socially beneficial. Ideas are created subject to an indivisibility. There are many possible ideas, and to model the fact that each should have downward sloping demand, we adopt the Dixit-Stiglitz model of preferences. The *private return* on an idea is the ratio of expected monopoly revenue to its cost of creation. We consider first the case in which the private return has a neutral effect on the relationship between the private and social benefit of an idea. We show that the complex heterogeneous mass of ideas can be analyzed by examining the total monopoly revenue from all ideas with a private return above a threshold level ρ. We refer to this function as $M(\rho)$. Using this tool, we show that when the market is sufficiently small, it may be optimal to provide an unlimited monopoly, but when the market is large enough, a time limit should always be imposed, and this limit should strictly decrease as the size of the market grows.

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 and Lai [2002, 2004]. 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. While this approach does not answer all dynamic issues, such as the depletion of existing ideas and ideas that use other ideas as inputs, it is a valid dynamic interpretation of the model we use. Since Grossman and Lai have already provided this interpretation, we do not do so here. 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. Although we do not think that this approach to the production of ideas is as useful as our disaggregated model of heterogeneous ideas, under the assumption of symmetry it is possible to translate production functions into equivalent total monopoly revenue functions, so this model has the same reduced form as ours, and we show how to make the connection.

The most significant difference between our results and those of Grossman and Lai are that they focus on the case in which the production function is Cobb-Douglas. Both they and we show that this means that optimal protection does not change with the size of the market. If, though, the total monopoly revenue function has increasing elasticity then optimal protection locally decreases and vice versa; the Cobb-Douglas case is then the boundary between these two general cases. Although we also show, as noted above, that when market size is large enough, optimal protection must always decrease - even in the Cobb-Douglas case - the question arises: are

we currently in a region of the total monopoly revenue function where optimal protection might be constant or increasing? We give both theoretical reasoning and empirical evidence that this is not the case.

Grossman and Lai focus also on "harmonization" and North-South trade. Because they provide a detailed treatment, especially of issues involving who benefits from harmonization, we examine only a simple example involving multiple countries. Our goal is to show how - contrary to their finding - in the empirically relevant case of increasing elasticity of total monopoly revenue, the North should reduce protection as a result of harmonization. In the case of two countries of equal size, because some of the benefits of a higher time limit are received by the other country there is a tendency to set protection too low, and there is a "harmonization" argument to be made for international treaties raising the time limit. However, this argument applies only to two countries of equal size. When the countries, two or more, are of unequal size, smaller countries tends to set low limits and free ride off of the large country - but the large country tends to set limits that are too high because it does not account for the social benefit of innovation to the smaller countries. An implication of this result is that the process of trade expansion should be accompanied by a parallel process of intellectual property reduction. In this case "harmonization" does not mean setting limits equal to or higher than those in the larger and more protected country, but rather adjusting the time limits to lie in between the larger protection of the larger country and the smaller protection of the smaller countries.

We emphasize several other points. Increasing the scale of the market can increase the demand for the specialized labor needed to create ideas. For given amount of intellectual property protection, this serves to drive up the wages of that kind of labor. We show that these increased economic rents do not serve a useful economic purpose, because they do not increase the number of ideas that are produced. Notice that this argument is relevant, at least in principle, for understanding the increased wage gap between skilled and unskilled workers. To the extent that activities with high IP protection employ a large share of skilled workers, an increase in the size of the market increases the rents accruing to the skilled workers, at least as long as the supply of the latter does not increase in proportion with the market.

The general equilibrium approach emphasizes the connection between broader features of the economy and intellectual property. We illustrate this through various comparative static results. We also consider a number of extensions of the basic model. The most important is to relax the assumption that quality is neutral in the relation between private and social values of an idea. As either the scale of the market or the term of intellectual property protection increases, ideas that are more marginal from a private point of view are produced. If the social value of these ideas declines even

faster, then the argument for decreasing the length of protection with market scale is strengthened. If ideas that are more marginal from a private perspective are more beneficial from a social perspective, then a system of exclusive rights is a poor method of encouraging the production of valuable ideas: it leads to the production of the least, rather than the most, socially useful ideas. When legislation is in place that allocates monopoly power to the producers of certain goods and not of others (patent and copyright protection vary widely across economic sectors) a natural consequence is the emergence of socially wasteful lobbying and rent-seeking activities. In our framework rent-seeking would consist of payments going from producers to legislators to increase the length of intellectual property protection. The basic result is that the entrepreneurs who have access to the less marginal projects are more willing to pay for the legal monopoly to be continued or extended. Hence, in presence of rent-seeking behavior, intellectual property protection leads to an equilibrium in which its length is determined, at the margin, by those innovators who need it the least.

One may wonder why the relation between the size of the market and the optimal level of IP protection (that is, the optimal degree of monopoly power) is not emphasized in the "new" growth literature, which has at its core a model of adoption of new commodities similar to the one used here. There are two answers: first, because the question is never explicitly asked as that literature assumes that monopoly power and IP protection are good, in fact: necessary, for innovation. Second, because the impact of market size on the optimal length of IP protection is buried behind the so called "size effect," according to which increasing the size of the market speeds up the growth process because of increasing returns. So, for example, in the Romer [1990] paper on endogenous technological change one reads that (p. S95) "... these models have an underlying form of increasing returns in research. As a result, an increase in a scale variable induces an increase in the rate of growth."

2. The Model

Ideas are indexed by their characteristics $\omega \in \Omega$. The space of characteristics Ω is a compact subset of a topological measure space. Each idea requires a minimum amount $h(\omega) \geq 0$ of the only primary input, labor, to be invented, where $h(\omega)$ is a measurable function. We refer to $h(\omega)$ as the *indivisibility* or the minimum size for producing a new idea. There can be many ideas with given characteristics; we describe this by a positive measure $\eta(\omega)$ representing the"number" of ideas with characteristics ω in an

economy of unit size. We will later focus on the case where $\eta(\omega)$ is a probability distribution, and innovators find their individual ideas by drawing from this underlying distribution, but this interpretation is not essential.

There is also a continuum population of agents of size λ; with the parameter λ we measure the scale of the economy. The number of available ideas may depend on the size of the economy, so the total number of ideas with characteristics ω available in an economy of size λ is $g(\lambda)\eta(\omega)$. To capture the principle that in a larger population more ideas of a given quality are available $g(\lambda)$ is assumed non-decreasing in λ; we may assume without loss of generality that $g(1) = 1$. In the case in which $\eta(\omega)$ is a probability distribution from which innovators draw their ideas, twice as many innovators mean twice as many ideas with given characteristics, so $g(\lambda) = \lambda$. Neither that the number of ideas increases with size at different rates for different characteristics, nor that the indivisibility varies with the size of the economy, is a possibility considered here.

We assume initially that once an idea is created, it may be reproduced at no cost and without limit. If the input of labor $y(\omega)$ used in producing ω is below the threshold, i.e. $y(\omega) < h(\omega)$ no prototype will emerge and no consumption is possible. If the amount of labor used in producing the idea exceeds the indivisibility, i.e. $y(\omega) \geq h(\omega)$, then consumption $x(\omega) \geq 0$. It is convenient also to measure consumption per capita as $z(\omega) = x(\omega)/\lambda$.

The utility of a representative individual has a Dixit-Stiglitz form over goods of different characteristics. When consuming z units of a good with characteristics ω, consumers receive a utility of $v(z,\omega)$, where $v(z,\omega) \geq 0$ is assumed continuous in ω and in z, non-decreasing in the latter, and at least up to a limit z^*, smooth and strictly increasing. We also assume $\lim_{z\to\infty} v(z,\omega) = v^C(\omega) < \infty$ and, to simplify the analysis of qualities that are not produced, $v(0,\omega) = 0$. Since $v(z,\omega)$ is bounded, $zv_z(z,\omega) \to 0$ as $z \to \infty$, that is per capita revenue falls to zero as per capita consumption grows without bound. We also assume that $zv_z(z,\omega)$ has a unique maximum at $z^M(\omega)$. Note that we are assuming that the utility from each characteristic is certain, or what amounts to the same thing, that v measures expected utility. Apart from consumption of idea-goods, consumers receive a utility of ℓ from leisure $0 \leq \ell \leq L$, where L is the individual endowment of time; L has two alternative uses: leisure or production. Leisure here is a shorthand for all activities that take place outside of the idea sector, so it includes the production of goods that do not involve ideas, or goods that use ideas that already exist. Since $g(\lambda)\eta(\omega)$ of type ω ideas are potentially available, individual utility is

$$\int v(z(\omega),\omega)g(\lambda)\eta(d\omega) + \ell.$$

Note that the marginal utility of leisure is normalized to one. We will consider later the possibility that the marginal utility of leisure changes with the scale of the economy, for example increases in per capita GDP may increase the productivity of labor in the non-idea sector. The social feasibility constraint is that the amount of leisure consumed equals the amount left over after the production of prototypes.

$$\lambda(L-\ell) = \int y(\omega)g(\lambda)\eta(d\omega).$$

Profit maximization and efficiency both require $y(\omega) = h(\omega)$ for all ideas for which $x(\omega) > 0$, and $y(\omega) = 0$, otherwise. It is also obvious that, as a consequence of the extreme form of the production technology, no good would be produced in this economy absent patent protection; we look later at the less extreme case in which a capacity constraint exists and making copies involves a positive marginal cost.

Patent Equilibrium. 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. We refer to ϕ as the level or the extent of protection. We assume that there are potentially many individuals who can invent any particular idea; certainly the number of individuals who have historically had truly unique ideas is minuscule. We do not explicitly model the "patent race" by which patent is awarded, and simply assume that through some procedure, a particular individual is awarded a "patent" for ideas with characteristics ω. This provides her with monopoly over those $\eta(\omega)$ ideas and, while the patent lasts, she chooses how many copies should be taken to the market to maximize profits. Under patent protection, our economy is similar to the traditional Dixit-Stiglitz "monopolistic competition" economy. Once a patent expires, anyone who wishes to do so may make copies of ideas that had been previously introduced under the patent regime. Once competition sets in, output and consumption jump to infinity while prices and revenue fall to zero. The latter we have assumed for the sake of simplicity; as pointed out in Boldrin and Levine [1999], this need not be, and it is not in general, the case. A type of good is produced if, given the patent length ϕ, the prospective monopolist finds it profitable to overcome the indivisibility. This notion of equilibrium is closely connected to that of Hart [1979] and Makowski [1980], and it has been used, for example, by Acemoglu and Zilibotti [1996] in a related context.

The market for innovation is equilibrated through the wage rate of labor w. The higher is w the costlier it is to produce new ideas and fewer of them will therefore be produced. If the amount of labor used in the production of

ideas is strictly less than the total endowment λL, wages $w = 1$. Otherwise, w must be chosen to reduce demand for labor to the point where the amount of leisure is 0.

A patent holder, monopolist, for a good with characteristics ω who sells z units of output to each of the λ consumers receives the price $v_z(z,\omega)$. Recall that revenue $\lambda z(\omega)v_z(z(\omega),\omega)$ is assumed to have a unique maximum at $z^M(\omega)$, and the cost faced by the monopolist is $wh(\omega)$. For a commodity with characteristics ω, $\rho(\omega) = z^M(\omega)v_z(z^M(\omega),\omega)/h(\omega)$ expresses the ratio of (per capita) private value to the innovation cost. In fact $\rho(\omega)/w$ represents one plus the rate of return on investment which would accrue to the inventor of commodity ω if patents lasted forever and the market size was $\lambda = 1$. We refer to $\rho(\omega)$ as the *private return* for ω. The monopolist receives a fraction ϕ of the private return, times the size λ of the market. Hence, a good is produced if

$$\rho(\omega) \geq w/\phi\lambda \equiv \underline{\rho}.$$

In other words, no ideas with private return lower than $\underline{\rho}$ will be introduced in the patent equilibrium and all ideas with a $\rho(\omega)$ above $\underline{\rho}$ will be produced. Notice that $\underline{\rho}$ is strictly decreasing in $\phi\lambda$, meaning that as the scale of the market or the extent of protection increase, ideas with a lower private return are introduced. Notice also that, in general, there need not be any monotone relation between the private return $\rho(\omega)$ of an idea and its social return; hence ideas of high social return may be introduced only for high values of λ, or even never at all, if their private return $\rho(\omega)$ is particularly low.

Per capita social welfare in a patent equilibrium is derived by integrating utility for those goods that are produced less the cost of producing them

$$\int_{\rho(\omega)\geq\underline{\rho}} [\phi v(z^M(\omega),\omega) + (1-\phi)v^C(\omega) - h(\omega)/\lambda]g(\lambda)\eta(d\omega) + L$$

We assume that for $\underline{\rho} > 0$

$$L^D = \int_{\rho(\omega)\geq\underline{\rho}}^{\infty} g(\lambda)h(\omega)\eta(d\omega) < \infty,$$

so that the amount of labor required to produce all ideas exceeding any particular private value threshold is finite.

Notice that $\rho(\omega)h(\omega)\eta(\omega)$ is the total revenue of a monopolist investing in goods with characteristics ω in an economy of unit size. For any given cutoff ρ we may define

$$M(\rho) = \int_{\rho(\omega)\geq\rho}^{\infty} \rho(\omega)h(\omega)\eta(d\omega).$$

Then $M(\rho)$ is the sum of monopoly revenue over all ideas with private value of ρ, or greater. 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.

Let $\nu^M(\omega) \equiv v(z^M(\omega),\omega)/[h(\omega)\rho(\omega)]$ and $\nu^C(\omega) \equiv v^C(\omega)/[h(\omega)\rho(\omega)]$ be the ratio of social value to private return of a commodity of type ω under monopoly and under competition, respectively. To fix ideas, consider the case in which utility has the quadratic form

$$v(\omega,z) = b(\omega)\left(Z(\omega)^2 - [z - Z(\omega)]^2\right)$$

for $z \leq Z(\omega)$ and $v(\omega,z) = b(\omega)Z(\omega)^2$ for $z > Z(\omega)$. Then we have $\nu^M(\omega) = 3/2$ and $\nu^C(\omega) = 2$ independently of characteristics. More generally, we can define the notion of *return neutrality*. If the ratios of social values to private return $\nu^M(\omega)$ and $\nu^C(\omega)$ are both constant, we have *strong return neutrality*. Formally, observe that the measure $h(\omega)\eta(\omega)$ represents, in an economy of unit size, the quantity of labor needed to produce all ideas with characteristics ω. Consider the measure $h(\omega)\eta(\omega)$, which we refer to as the *mass of ideas*, restricted to the σ-subalgebra of the Borel sets of Ω generated by the subsets of Ω on which $\rho(\omega)$ is constant; make the regularity assumption that it can be represented by a continuous density function $\mu(\rho) = \int_{\rho(\omega)=\rho} h(\omega)\eta(d\omega)$. For any function $f(\omega)$ we may define a conditional value $\overline{f}(\rho)$ in much the same way as a conditional expectation is defined. Specifically, $\overline{f}(\rho)$ is defined, μ-almost everywhere, by the condition that $\int_B \overline{f}(\rho)\mu(\rho)d\rho = \int_B f(\omega)h(\omega)\eta(d\omega)$ for every set B in the σ-subalgebra of the Borel sets of Ω on which $\rho(\omega)$ is constant. By *return neutrality* we mean that $\overline{\nu}^M(\rho), \overline{\nu}^C(\rho)$ are constant.[1] Below, we consider first the neutral then the non-neutral case.

3. Return Neutrality

We first examine the case of return neutrality, and ask how socially optimal protection $\hat{\phi}(\lambda)$ depends on market size. We find that, if the elasticity of total monopoly revenue is well behaved near $\rho = 0$ then for large enough λ socially optimal protection must be declining with λ. Further, if the elasticity of total monopoly revenue is increasing with ρ, a condition that, contra Grossman and Lai [2002, 2004] we argue is likely to be the case, then socially optimal protection is in fact decreasing as a function of λ. Basically, there are two cases. If the elasticity of total monopoly revenue is increasing with ρ and $\hat{\phi}(\lambda) < 1$ we can show from the first order conditions and implicit function theorem that $\hat{\phi}(\lambda)$ is strictly decreasing. If, instead, the elasticity of total monopoly revenue is decreasing with ρ, then labor demand is growing

[1] By assuming all ideas are identical from the point of view of consumers Grossman and Lai [2004] implicitly assume strong return neutrality.

faster than labor supply, and so the labor constraint must eventually bind. We show that whenever the labor constraint binds, it must be the case that $\hat{\phi}(\lambda)$ is strictly decreasing.

Proposition. *Suppose return neutrality. If for some* $\tilde{\rho}$ *and* $0 < \rho < \tilde{\rho}$, $\Upsilon'(\rho) \neq 0$ *then there exists* $\overline{\lambda}$ *such that* $\hat{\phi}(\lambda)$ *is unique and strictly decreasing for* $\lambda > \overline{\lambda}$. *If* $\Upsilon'(1/\lambda\widehat{\phi}(\lambda)) > 0$ *then* $\hat{\phi}(\lambda)$ *is unique and non-increasing; if* $\hat{\phi}(\lambda) < 1$ *it is strictly decreasing.*

Proof. Use return neutrality to rewrite social welfare as

$$\int_{\rho' \geq \underline{\rho}} [\phi\overline{v}^M \rho' + (1-\phi)\overline{v}^C \rho' - 1/\lambda] g(\lambda)\mu(\rho')d\rho' + L$$

We begin by analyzing the case in which the labor constraint does not bind, so $w = 1$. Differentiating with respect to ϕ and dividing out the constant $g(\lambda)$ we get the first order condition for a social optimum.

$$\begin{aligned} FOC(\lambda,\phi) = & \\ & \left[(1/\phi)\left\{\phi\overline{v}^M + (1-\phi)\overline{v}^C\right\} - 1\right](1/\lambda^2\phi^2)\mu(1/\phi\lambda) \\ & - \int_{1/\phi\lambda}^{\infty} \rho(\overline{v}^C - \overline{v}^M)\mu(\rho)d\rho \\ & = -\left[(1/\phi)\left\{\phi\overline{v}^M + (1-\phi)\overline{v}^C\right\} - 1\right](1/\lambda\phi)M'(1/\phi\lambda) \\ & - (\overline{v}^C - \overline{v}^M)M(1/\phi\lambda) \end{aligned}$$

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

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

has the same qualitative properties as $FOC(\lambda,\phi)$: it has the same zeroes, 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

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

The second terms is unambiguously negative. The first term has two factors of interest. We have $(1/\phi)\left\{\phi\overline{v}^M + (1-\phi)\overline{v}^C\right\} - 1$ representing social surplus of the marginal idea produced; since privately it yields zero

profit, it must yield positive social surplus. If the other factor $\Upsilon'(1/\lambda\phi) > 0$ 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

$$\frac{d\phi}{d\lambda} = -\frac{NOC_\lambda}{NOC_\phi} \propto NOC_\lambda =$$

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

which is negative if and only if $\Upsilon'(1/\lambda\phi) > 0$. This covers the second half of the proposition when the labor constraint does not bind.

If the labor constraint does bind, increasing ϕ only increases the wage rate. Hence, if the social optimum is to allow the labor constraint to bind, ϕ must be chosen as small as possible subject to the constraint of full labor utilization and $w = 1$. Consequently concavity of welfare in the interior implies a unique optimal choice of ϕ. This establishes a unique optimal policy function $\hat{\phi}(\lambda)$.

Finally, we turn to the first half of the proposition. For fixed ϕ and all large enough λ we may assume that either $\Upsilon'(1/\lambda\phi) > 0$ or $\Upsilon'(1/\lambda\phi) < 0$. In either case, $\Upsilon(\rho)$ must have a (possibly infinite) limit as $\rho \to 0$. Observe that $\Upsilon(\rho) \equiv -\rho M'(\rho)/M(\rho)$, and that $M(\rho)$ is non-increasing. Suppose first that $-\rho M'(\rho)$ does not converge to infinity. If it is bounded away from zero, $M(\rho) \to \infty$, implying $\Upsilon(0) = 0$. If it is not bounded away from zero, since $M(\rho)$ is bounded away from zero, again, $\Upsilon(0) = 0$. Hence, either $-\rho M'(\rho) \to \infty$ or $\Upsilon(\rho) = 0$. The latter case implies $\Upsilon'(1/\lambda\phi) > 0$, so fix $\phi = 1$ and examine

$$NOC(\lambda,1) = [\sigma - 1]\,\Upsilon(1/\lambda) - (\overline{v}^C - \overline{v}^M).$$

Since $\Upsilon(0) = 0$ for λ sufficiently large $NOC(\lambda,1) < 0$ implying $\hat{\phi}(\lambda) < 1$.

Finally, then, suppose $-\rho M'(\rho) \to \infty$. The demand for labor is

$$L^D = g(\lambda)\int_{1/\phi\lambda}^{\infty} \mu(\rho)d\rho.$$

Differentiating with respect to λ yields

$$D_\lambda L^D = g'(\lambda)\int_{1/\phi\lambda}^{\infty} \mu(\rho)d\rho + g(\lambda)(1/\phi\lambda^2)\mu(1/\phi\lambda).$$

Labor supply is λL so if $D_\lambda L^D \geq L + \varepsilon$ for all sufficiently large λ the labor constraint must eventually bind. But $-\rho M'(\rho) = \rho^2\mu(\rho) \to \infty$ as $\rho \to 0$ so, for ϕ bounded away from zero, $D_\lambda L^D \to \infty$, so in this case the labor constraint must bind. □

Three Implications. Theorists of innovation and growth due to aggregate increasing returns often insist that the latter are due to massive externalities in the creation of new ideas. In our framework this requires $g(\lambda)/\lambda$ to be increasing in λ, that is the number of available ideas increases more rapidly than the scale of the economy. This leads to a number of extravagant predictions, which we will compare to the empirical evidence later. Here it suffices to remark that, as a matter of theory, $g(\lambda)/\lambda$ increasing in λ implies that the labor constraint must bind eventually; when it does the private return from the marginal idea adopted increases rather than decreases. While this does not imply that ideas with a higher social value will be produced, it certainly does imply that the optimal length of protection ϕ is strictly decreasing in λ.

That intellectual property protection can drive up the wage rate for the relevant supply of highly skilled labor, is empirically relevant for growing and advanced economies alike. Lobbyist groups, especially in the entertainment and pharmaceutical industries, often point to the high cost of producing new goods as a reason for strong IP protection. Examination of the balance sheets of either movie production or pharmaceutical companies shows that much of this high cost is due to the earnings of the "highly skilled" labor these, and other IP protected industries, employ. Consider the entertainment industry: costs are high here because a few "stars" earn large amounts of money, Since the opportunity cost of these people is generally quite small an important effect of reducing copyright protection will simply be to lower the rents earned by these "stars," and consequently reduce the cost of producing movies of a given quality. In such industries, the marginal workers are paid close to their opportunity cost, and so stand to lose little through reduced copyright protection. A similar, even if admittedly less straightforward, argument can be applied to the drug industry with respect to the wages of medical researchers in relation to that of production workers.

Other, policy relevant, comparative static results also follow from the previous analysis. Any policy or technological change increasing the marginal cost of the skilled labor needed to introduce new ideas is equivalent to decreasing λ, so generally such changes increase the socially optimal length of protection. This observation is relevant when confronting policies that increase the protection from foreign labor competition (for example, by restricting immigration, or penalizing outsourcing) of selected groups of skilled workers; in this case, protection in the labor market induces additional intellectual property protection and rent-seeking in the product market. Conversely, technological improvements (such as the increasing power and reduced cost of computers, or the reconstruction of the DNA code)

which reduce the size of the initial indivisibility h are tantamount to increasing λ, hence should engender a reduction of the socially optimal length of protection.

The Production Function Approach. By way of contrast, Grossman and Lai [2002, 2004] adopt a production function approach where ideas are homogenous and the total number of ideas is $Q = F(H,L)$ where H is a fixed amount of human capital, L is the labor input and F is a constant return to scale concave production function. Since human capital serves in this model only to absorb the rents from ideas, we may as well write $Q = f(L)$ where f is a diminishing return production function. While we do not think that ideas are like automobiles - more or less perfect substitutes stamped off a production line - we can construct a total monopoly revenue schedule corresponding to Grossman and Lai's production function, and by doing so give an interpretation of their assumption in our framework. Observe that when $w = 1$ the total labor cost of producing Q ideas is $f^{-1}(Q)$, and the corresponding marginal cost is $1/f'(L)$. Since all ideas are equally valuable, we may as well suppose they generate revenue 1, so in our terminology, the private return of the idea produced by the Lth unit of labor input is revenue divided by the cost of producing the idea, that is, $\rho = f'(L)$. The total revenue to ideas with private return ρ or better is then the total number of ideas produced by the corresponding amount of labor

$$M(\rho) = f([f']^{-1}(\rho)).$$

From this we may easily derive that the elasticity Υ of M is the same as Grossman and Lai's elasticity of research output with respect to labor.[2] In the Cobb-Douglas case, which is the benchmark case studied by Grossman and Lai, $f(L) = L^{\alpha}$, and so $M(\rho) = \alpha^{\alpha/(1-\alpha)}\rho^{-\alpha/(1-\alpha)}$, which is to say that M has constant elasticity. Also of interest is their CES case, where $f(L) = (a + L^{\beta})^{1/\beta}$for $\beta \leq 1$. We have $f'(L) = (aL^{-\beta} + 1)^{(1-\beta)/\beta}$,

$$M(\rho) = a^{1/\beta}\left(\frac{1}{1-\rho^{-\beta/(1-\beta)}}\right)^{1/\beta}$$

$$\Upsilon(\rho) = \frac{1}{1-\beta}\frac{1}{\left(\rho^{\beta/(1-\beta)}-1\right)}.$$

For $\beta < 0$, $M(\rho)$ is defined for $\rho \leq 1$, with $M(1) = 0$, and elasticity is increasing in ρ. The case $0 < \beta \leq 1$ implies that even with no labor input, $a^{1/\beta}$ ideas will be produced. This is not ridiculous: even in the absence of any effort, some ideas with a positive return may be discovered by accident. Also in this case $M(\rho)$ is defined for $\rho > 1$ with $M(1) = \infty$, that is, the

[2] In their notation, this elasticity is γ.

revenue generated by ideas goes to infinity, even before all ideas of positive quality are exhausted. Here we have decreasing elasticity.

4. The Elasticity of Total Monopoly Revenue

The question then is, which assumption is most appropriate on the elasticity of the total monopoly revenue function? We address it here from both a theoretical and empirical perspective. First we look at distributions that are interesting either because they satisfy intuitive properties or because they allow for explicit computations and are widely used in practical examples. With the exception of the Pareto, we find that the elasticity of $M(\rho)$ is increasing. Next we look at empirical distributions of income/revenue for different kinds of inventive activity, and find that they exhibit rapidly rising elasticity. Finally, we derive the implications of the constant elasticity assumption for the equilibrium demand of labor in the idea sector, as the scale of the market increases; we show that assumptions of constant or decreasing elasticity is grossly at odds with any available data, and probably also with common sense.

4.1. **Theoretical Analysis.** First, we consider from a theoretical point of view what

$$M(\rho) = \int_{\rho}^{\infty} \rho' \mu(\rho') d\rho',$$

might look like. Under the plausible assumption that there are ideas so bad that they have a negative private return we expect $\mu(0)$ to be strictly positive, and finite. This implies that $M(0)$ is finite, and $M'(0) = 0$, so $\lim_{\rho \to 0} \Upsilon(\rho) = 0$. Since $\Upsilon(\rho) \geq 0$, this means $\Upsilon'(0) \geq 0$, that is the increasing elasticity case. In other words, theoretical considerations alone suggest that the function $M(\rho)$ is finite and flat at $\rho = 0$, and has increasing elasticity there; we should not expect situations such as that implied by the CES production function for $0 < \beta \leq 1$ in which elasticity is globally decreasing, or even constant.

4.1.1. *The Pareto Distribution.* If $\mu(\rho)$ is Pareto, then $M(\rho) = \rho^{-\zeta}$, which corresponds, up to a scale factor, to the functional form implied by the Cobb-Douglas production function. Since the Pareto density goes to infinity for finite ρ, as we have observed, this is not a terribly good global model of the distribution of ideas, and we would not expect it to hold for ρ close to zero. Nevertheless, it can certainly be argued that we are still in the upper tail of the distribution of quality of ideas. And there is certainly a great deal of economic data that appears to have Pareto upper tails, so it is possible in principle that we are still experiencing values of ρ for which $\Upsilon'(\rho) \leq 0$,. Or, perhaps, the tails are even thicker than Pareto, with a lump

of ideas with zero cost, as is the case for the CES production function with $0 < \beta \leq 1$. This is an empirical issue, and we will address it shortly.

Next we quickly run through various functional forms for $\mu(\rho)$ that might be useful to model the distribution of ideas, and see what they imply for $M(\rho)$, and $\Upsilon(\rho)$.

4.1.2. *The Exponential, Normal and Lognormal Distributions.* Consider first the Exponential, where $\mu(\rho) = \zeta^3 e^{-\zeta\rho}$. Then $M(\rho) = \zeta(\zeta\rho + 1)e^{-\zeta\rho}$ and

$$\Upsilon(\rho) = (\zeta\rho)^2/(1+\zeta\rho)$$

which is increasing in ρ. For the Normal distribution, say $\mu(\rho) = e^{-\zeta\rho^2}$, we have that $M(\rho) = (2\zeta)^{-1}e^{-\zeta\rho^2}$ and $\Upsilon(\rho) = 2\zeta\rho^2$, which is also increasing for all values of $\rho > 0$. Similar calculations yield the same results for a Lognormal distribution of private returns. Should we restrict ourselves, as it it reasonable, to positive and large values of ρ the results do not change, in fact they are strengthened. For the Normal distribution, $\Upsilon(\rho)$ is increasing in ρ over the whole interval $(0,\infty)$, constant at $\rho = 0$ and decreasing for negative values of ρ. For the Exponential, $\Upsilon(\rho)$ is increasing for all $\rho > 0$ and for $\rho < -2/\zeta$ as well.

4.1.3. *The Truncated Pareto Distribution.* In principle, a Pareto distribution with a finite upper bound is appealing, and it allows for explicit computations. It amounts to assume that private returns are distributed Pareto and there is an idea of highest possible quality, so $\mu(\rho) = \rho^{-\zeta}$ for $\rho \leq \overline{\rho}$, and $\mu(\rho) = 0$ for $\rho > \overline{\rho}$, with $\zeta < 2$ (but notice that for values of $\zeta < 0$ this says that the frequency of ideas increases with their quality, and then suddenly drops to zero after the maximum value $\overline{\rho}$ is reached, which does not make much practical sense. so assume $\zeta > 0$.) In this case, again, elasticity rather than being constant is increasing. In fact, we have

$$M(\rho) = \frac{1}{2-\zeta}\left[\overline{\rho}^{2-\zeta} - \rho^{2-\zeta}\right]$$

$$\Upsilon(\rho) = \frac{2-\zeta}{(\overline{\rho}/\rho)^{2-\zeta} - 1}$$

$$\Upsilon'(\rho) = \frac{(2-\zeta)^2\,(\overline{\rho})^{2-\zeta}\rho^{\zeta-3}}{\left((\overline{\rho}/\rho)^{2-\zeta} - 1\right)^2} > 0.$$

In the case $\zeta = 1$ we see that $M(\rho)$ is linear, and we can explicitly solve the NOC

$$NOC(\lambda,\phi) = \left[(1/\phi)\left\{\phi\overline{\nu}^M + (1-\phi)\overline{\nu}^C\right\} - 1\right]\frac{1}{\lambda\phi\overline{\rho} - 1} - (\overline{\nu}^C - \overline{\nu}^M) = 0.$$

We rewrite the NOC as

$$\lambda\overline{\rho}(\overline{v}^C - \overline{v}^M)\phi^2 + \phi - \overline{v}^C = 0$$

and then find the positive root

$$\phi = \frac{\sqrt{1 + 4\lambda\overline{\rho}(\overline{v}^C - \overline{v}^M)\overline{v}^C} - 1}{2\lambda\overline{\rho}(\overline{v}^C - \overline{v}^M)}.$$

If we normalize $\overline{\rho} = 1$, then for large λ this can be approximated by

$$\phi \approx \left(\frac{\overline{v}^M}{\overline{v}^C - \overline{v}^M}\right)^{1/2} \lambda^{-1/2}$$

which implies an elasticity of protection with respect to scale of market of 1/2. That is, quadrupling the scale of the market implies that protection should be reduced by about a factor of two.

Returning to the case of general ζ, consider the NOC for $\phi = 1$.

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

If $\lambda > 1/\overline{\rho}$ holds, that is, the size of the economy is large enough, then $NOC(\lambda, 1) < 0$ implying an interior solution. This shows that the size of the economy and the upper bound on the achievable private return on ideas play a similar role in our model; a small economy where very highly profitable ideas are available is equivalent, from the viewpoint of optimal protection, to a large one in which only not-so-profitable ideas are available.

Now, let $\overline{\rho}\lambda \leq 1$ hold, so that $NOC(\lambda, 1) > 0$, so $\phi = 1$ is optimal. This is an economy either of very small size or in which the private return from new ideas is very low. In this case even complete monopoly cannot help: with $\phi = 1$ the marginal idea produced is $\rho = 1/\lambda$ implying that $\rho \geq \overline{\rho}$ holds, that is, no idea is ever implemented.

Finally, notice that, in general, labor demand is

$$L^D = g(\lambda) \int_{1/\lambda\phi}^{\overline{\rho}} \rho^{-\zeta} d\rho = \frac{g(\lambda)}{\zeta - 1}[(\phi\lambda)^{\zeta-1} - \overline{\rho}^{\zeta-1}]$$

while labor supply is $L^S = L\lambda$. Suppose the labor constraint is binding; equilibrium is achieved either by lowering labor demand through the wage rate, or through the protection level ϕ. We know social optimality requires keeping the wage rate at 1, and equilibrating the labor market using the protection level ϕ. This give the optimal protection level of

$$\phi = \left[\left(\frac{1}{\lambda\overline{\rho}}\right)^{\zeta-1} + \frac{(\zeta - 1)L}{g(\lambda)\lambda^{\zeta-2}}\right]^{1/\zeta-1}$$

which, once again, is strictly decreasing in λ and $\overline{\rho}$.

4.2. **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 her own ideas; that is, to think of ω as indexing individuals or, rather, individuals' private return from investing in ideas. In this case one can identify ω with ρ, and just use the latter. Then $h(\rho)$ becomes equal to the cost of person ρ's time, or if labor is equally productive in the "leisure" sector, $h(\rho)$ would be the same for all individuals. This is a strong, but not incredible assumption that is more or less the opposite of the production function approach; we will adopt it to start with and consider its weaknesses later. That is, set $h(\rho) = 1$; then, any individual with an idea of expected value higher than 1 would try to implement it. We then think of individuals 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" $\underline{\rho} = 1/\phi\lambda$.

An issue arises when going to the data because, for each individual ρ, we seldom observe the expected value of her idea, i.e. ρ itself, but, instead, one or more realizations $x_\rho = \rho + \varepsilon_\rho$, where ε_ρ is some uncontrollable source of randomness, possibly dependent on ρ itself. To the extent that the variance of ε_ρ is small relative to ρ, the sample distribution of x_ρ is an acceptable approximation for the underlying $\mu(\rho)$, and our empirical analysis makes sense. When (the variance of) ε_ρ is large, our analysis is invalidated, further still when the dependence on ρ is relevant; the impact of this limitation if discussed at the end of the section. Under the assumption that ε_ρ is small we now make several attempts at using available data to measure $M(\rho)$ directly, focusing especially on cases where it is likely that ε_ρ is small.

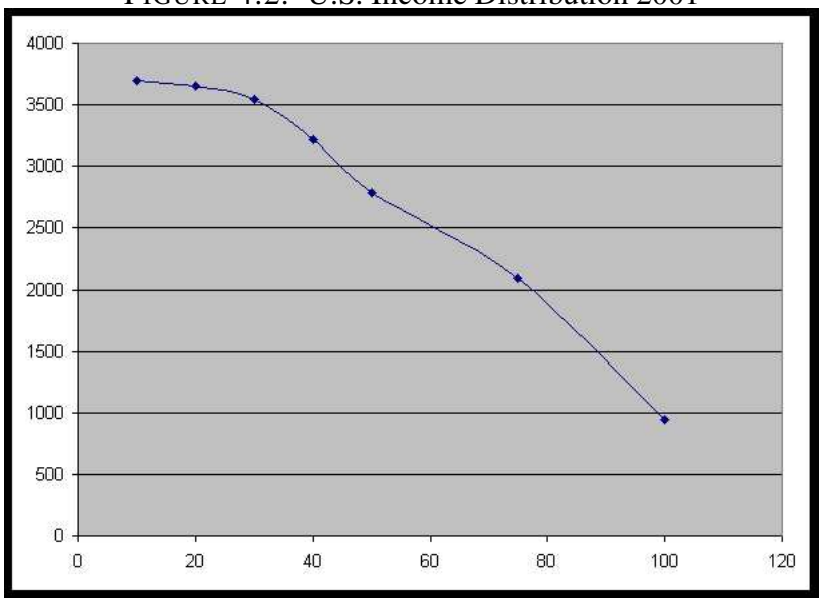
4.2.1. *Personal Income Distribution.* Our first attempt uses data for the US income distribution. That is, we make the further assumption that the distribution of income among creative individuals is the same as for the population at large. This is probably incorrect when it comes to levels: creative individuals are likely to be concentrated in the upper tail of the distribution of personal income. However, it is a plausible assumption about the shape of the distribution of income from creative activity. In any case, to the extent that the largest share of personal income is due to labor effort and people use their creativity in accumulating skills and choosing an occupation, this is a reasonable starting point. Current Population Survey Data on income from 2001 is shown in Figure 4.1. Abusing notation slightly, let ρ denote personal income here, the corresponding $M(\rho)$ is plotted in Figure 4.2.

FIGURE 4.1. U.S. Income Distribution 2001

Income in USD 1,000s	Population
0-10	9.0
10-20	6.9
20-30	13.3
30-40	12.4
40-50	15.4
50-75	18.4
75-100	10.8
100+	13.8

Source: U.S. Census Bureau, Historical Income Tables, Households, All Races http://www.census.gov/hhes/income/histinc/h17.html

FIGURE 4.2. U.S. Income Distribution 2001



Sophisticated econometric and statistical knowledge is not required to see that this curve is well fit by a straight line and poorly by a Pareto distribution. The US cumulative distribution of personal income, clearly, has increasing elasticity.

4.2.2. *Revenue from Authorship of Fiction Books.* We now examine a particular category of creative individuals: authors of fictional books. Ideally one would like to observe revenues for various draws of x_ρ (i.e. books) for each author ρ, to account for the possible impact of ε_ρ. Such data is not available, hence we proceed with what is. Although we do not have data

on life time income of individual authors, we do have data on the revenue generated by individual book sales. Looking directly at revenues from book sales is interesting because, as noticed, authors may earn above average income, and indeed may be largely in the upper tail of the income distribution - which is not measured at all by the income data we reported earlier. It might well be that while the bulk of the distribution of personal income is linear, it is Pareto in the upper tail, and the income of authors might turn out to be distributed like the upper tail of overall personal income. Assuming $h(\rho)$ constant in this case means that all those that "give it a try" at writing fiction have equivalent opportunity costs. It is hard to point to a measurable source of heterogeneity in the value of leisure among fiction writers *before* their success, or failure, is revealed through repeated sales of various books. *Ex-post*, obviously, successful writers are successful and their opportunity cost of time is high; but how predictable success and failure were when they decided to write their first couple of novels? For our data analysis to make sense one needs to assume that authors roughly know how much (each of) their books will sell. Suppose that

- All authors take the same amount of time to produce a novel and have the same opportunity cost so $h(\rho)$ is constant.
- Authors earn all their income from the sale of their novels.
- Expected revenues from the sale of a book are perfectly anticipated.

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 also ignore the fact that it is costly to produce books once they are written, which is largely irrelevant to our ends to the extent that the cost of producing each copy of a book is independent of the number of copies produced and sold. We also ignore that, for new authors, there is an option value to producing the first book - since if it is a failure, there is no reason to continue as an author. Our static model does not capture this type of option value. But if we assume that the cost of writing the first book is small relative to the lifetime cost of being an author, this will not matter much. With these caveats, we gathered data on revenues for 1223 and 1235 fiction books published in September 2003 and September 2004 respectively. Figure 4.3 shows $M(\rho)$ for September 2003. Figure 4.4 shows a plot on logarithmic axes, including a closeup do illustrate more clearly the increasing nature of the elasticity. on both ordinary and logarithmic axes. Figure 4.5 shows the September 2004 data,.

Examining these plots, four things stand out. First, they are very similar both in shape and absolute values; this suggests that (i) the pattern reported

FIGURE 4.3.

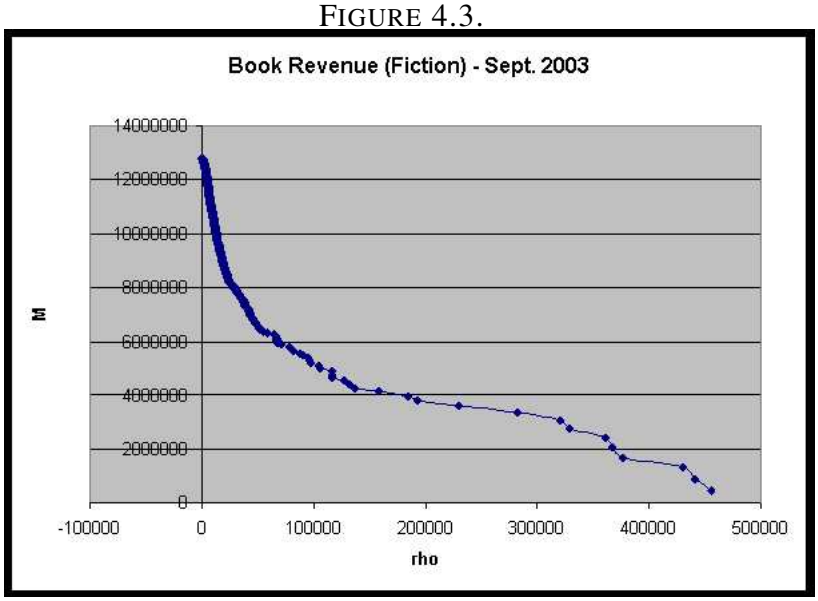
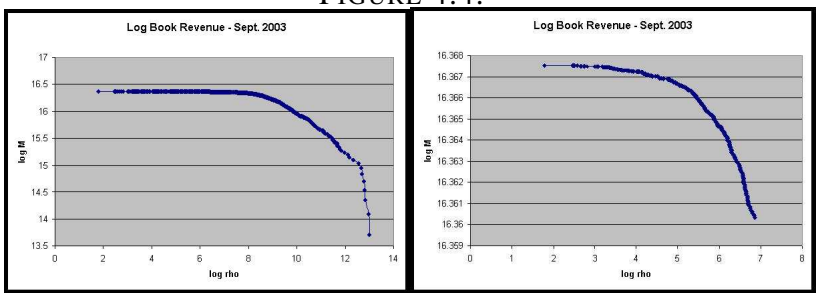


FIGURE 4.4.

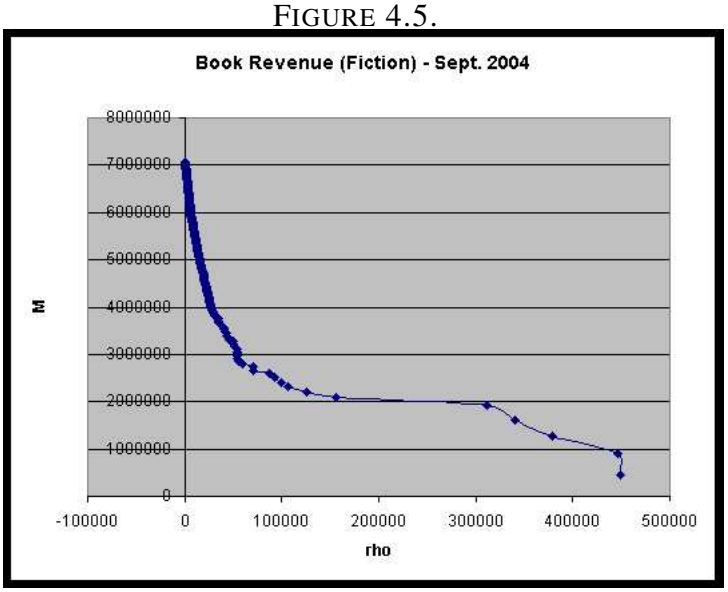


here is quite robust, and, (ii), even for successful books most sales take place in the first few months after publication.[3] Second, that the $M(\rho)$function is piecewise nearly linear and exhibits increasing elasticity - a fact that can be seen more clearly in the logarithmic plots. The third striking feature is how little sales between roughly \$150,000 and \$300,000 contribute to total revenue.[4] The fourth is how low the indivisibility $wh(\rho)$ may be for writing and publishing fiction; in 2003, 1181 books, out of a total of 1223, earned \$50,000 or less (corresponding to total revenue of approximately \$300,000.) These books accounted for 50 per cent of total revenue, that is \$6M out of \$12M. The numbers for 2004 are similar. In the September 2003 vintage,

[3]Data for the months of March 2003 and 2004 confirm this is indeed the case.

[4]The sales data are from a single distributor Ingram, constituting about 1/6th of the US book market, so total US revenues would be about six times this number.

FIGURE 4.5.



984 books earned less than $10,000, hence our estimate of the marginal author's opportunity cost $wh(\underline{\rho})$ should be placed at $60,000 or less. In the light of our earlier discussion, such conclusion is valid only if those authors, and their publishers, correctly predicted actual sales. Certainly books earning $300,000 in revenues are paying the opportunity costs of the authors, and it is hard to imagine that 1235 books were published during September 2004 in hopes of being one of the 5 that generate more than $300,000 in revenue.

4.2.3. *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*. In this case, by disaggregating by industry, it is at least plausible that the fixed cost of the innovation is not systematically related to the *ex post* private return. We use *ex post* data on the value of patents from Lanjouw [1993] for four German industries - estimated from patent renewal rates and data on the cost of renewal. We graph the corresponding $M(\rho)$curves in Figure 4.6.

We compute the elasticities of the linear spline at the midpoint of the intervals in Figure 4.7.

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. Moreover, with the (irrelevant) exception of the highest category of ρ for computers, they exhibit increasing elasticity over every portion of their range. Figure 4.7 also reports in square

FIGURE 4.6.

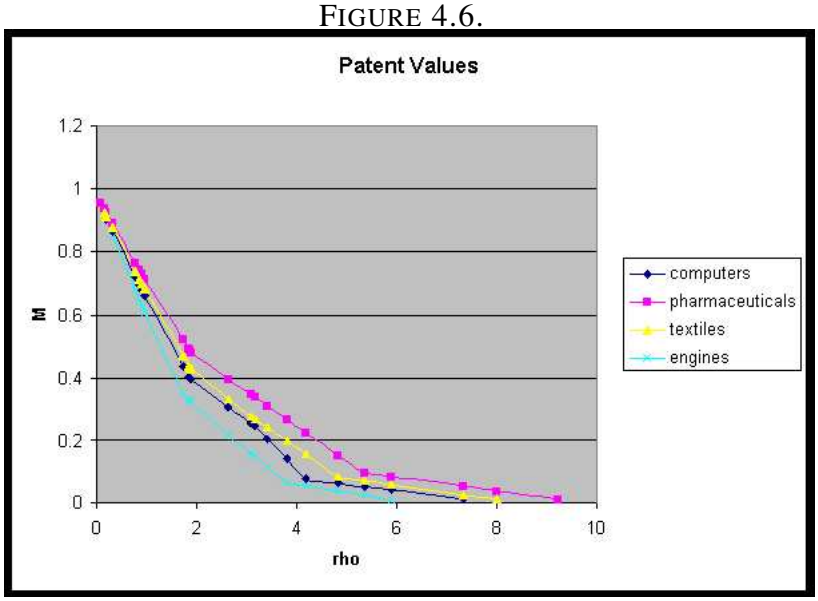


FIGURE 4.7. Elasticities

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]

brackets $-\rho M'(\rho)$. This is relevant because the same ϕ (i.e. 20 years) applies across sectors. Hence the relevant distribution is $M(\rho) = \sum_i M_i(\rho)$, where i indexes the various industries subject to patenting. Unfortunately, the fact that each $M_i(\rho)$ function has increasing elasticity does not imply that this is true in the aggregate. Hence it is of interest to examine $-\rho M_i'(\rho)$. If this is increasing, then the corresponding elasticity is increasing as well, and increasing $-\rho M_i'(\rho)$ is a condition that does aggregate. While not increasing in virtually every case as is the elasticity, $-\rho M_i'(\rho)$ is increasing at the lower end (near the existing threshold) and increases in most cases. This means that these results can be expected to aggregate over the different industries.

One thing that should be clear about this analysis is that existing data is not ideally suited to examining the $M(\rho)$ function. In particular, estimate of the anticipated and unanticipated components of return, and of the fixed cost $h(\omega)$ would greatly improve the analysis.

Our findings for patents appear to accord well with existing, and technically more sophisticated, empirical literature on the same subject. 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 performances of various empirical distributions, including Pareto's, to fit the data and find that a two-parameter log normal distribution provides the best fit. Silverberg and Verspagen (2004) use a variety of different data sources (European and US Patent Offices, as well as data set on CT scanners), and different measures of ρ (citations and monetary values) in their estimation. 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 lower cutoff value $\underline{\rho}$, this is supportive of our claim. Interestingly, Sampat and Ziedonis (2002) find that citations are not a good predictor of revenues earned from licensed patents. The older literature on the value of patents, stemming from the path-breaking paper of Pakes [1986] (see Hall, Jaffe, and Tratjenberg [2004] for a recent update and new results), seems to find, almost always that the appropriate distribution is a log-normal or an exponential, for both of which the elasticity of the total revenue function is increasing. Interestingly, Sampat and Ziedonis (2002) find that citations are not a good predictor of revenues earned from licensed patents.

4.2.4. *Analysis of Labor Demand: Theory.* Another route to determining whether elasticity is increasing or decreasing is to study its implications for the labor demand in the idea sector. This is

$$L^D(\lambda) = g(\lambda) \int_{\underline{\rho}=1/\phi\lambda}^{\infty} \mu(\rho) d\rho = g(\lambda)\ell(\underline{\rho}),$$

and letting El denote the elasticity operator, its elasticity is

$$El_\lambda[L^D(\lambda)] = El_\lambda[g(\lambda)] - El_{\underline{\rho}}[\ell(\underline{\rho})]$$

Depending on which assumptions one makes about $g(\lambda)$, the first factor ranges from zero to any large positive number. For example, Grossman and Lai [2004] identify $g(\lambda)$ with aggregate human capital H in their model, and assume this is constant relative to market size; hence $El_\lambda[g(\lambda)] = 0$. As pointed out in Section 3, in models of growth and innovation due to externalities, such as Grossman and Helpman [1991, 1994, 1995], or Romer [1990], $g(\lambda)$ increases faster than λ, hence $El_\lambda[g(\lambda)] > 1$. A benchmark case is that in which each individual draws her own ideas from the same urn, with or without replacement. If sampling is without replacement, and each person draws the same number of ideas for each characteristic ω, then

$g(\lambda) = \lambda$ and $El_\lambda[g(\lambda)] = 1$; if sampling is with replacement then, in general, we would have $El_\lambda[g(\lambda)] =< 1$.

As for the second factor, notice first that the demand for labor is

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

Now, assume that $\mu(\rho) = \rho^{-(\zeta)}$, with $\zeta > 2$ so $M(\rho) < \infty$, which is the constant elasticity case. Then

$$\ell(\underline{\rho}) = \frac{(\underline{\rho})^{1-\alpha}}{\alpha - 1}$$

and

$$El(L^D(\lambda)) = El(g(\lambda)) + \alpha - 1 > 1.$$

Notice that, when $El(g(\lambda)) > 0$ the elasticity of labor demand is predicted to be *substantially* larger than one. This implies that in the data we should observe that, as the size of the economy grows, the *share* of workers in the idea sector grows more than proportionally. This prediction is reinforced when the elasticity of the total monopoly revenue function is decreasing, as we show next.

Proposition. *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 $El[M\ell_1(\rho)] > El[\ell_2(\rho)]$.

(3) As the elasticity of total revenue goes from increasing, to constant, to decreasing, the elasticity of the associated labor demand functions increase 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 $D^2M_2(\rho) - D^2M_1(\rho) = (M(\rho)/\rho)[D\Upsilon_1(\rho) - D\Upsilon_2(\rho)] < 0$, 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: $El[\ell_1(\rho)] > El[\ell_2(\rho)]$

Because

$$El[\ell(\rho)] = El[\int_\rho^\infty -[DM(\rho')/\rho']d\rho'] =$$

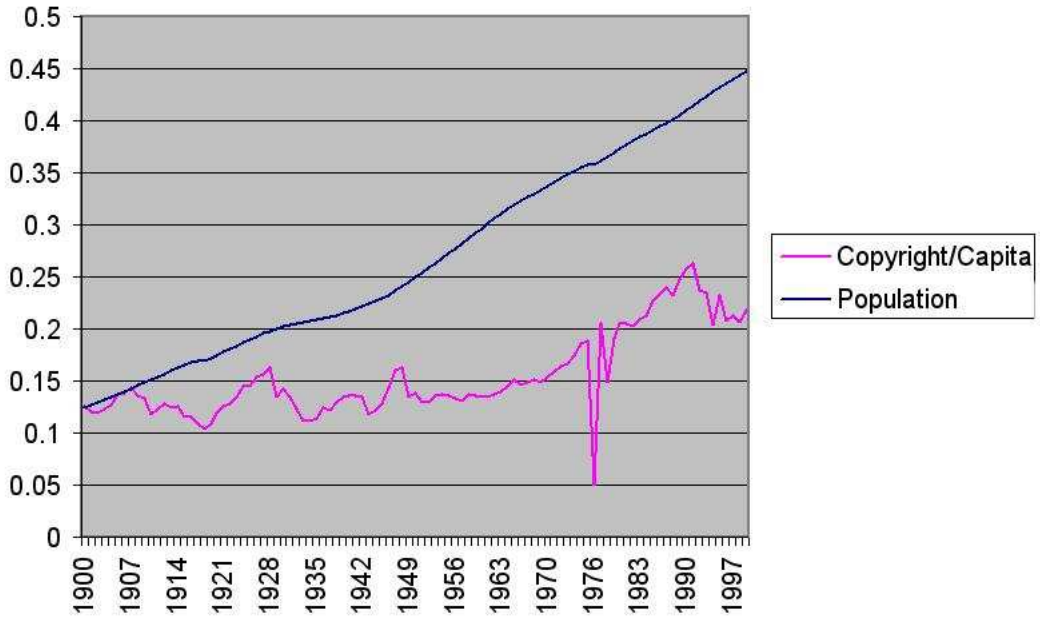
$$= \frac{-\rho DM(\rho)/\rho}{\int_\rho^\infty -[DM(\rho')/\rho']d\rho'} =$$

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

$El[\ell_1(\rho)]$ and $El[\ell_2(\rho)]$ have the same numerator, and, because of Step 2, the first has a smaller denominator. Hence the conclusion. □

In plain words: if a constant elasticity revenue function implies an elasticity of labor supply respect to market size larger than one, then a revenue function with decreasing elasticity would imply an even larger elasticity of labor supply. Playing this backward: should the empirical elasticity of labor supply with respect to changes in market size be clearly smaller than one, then the associated total revenue function would have to have increasing elasticity, which is our claim. Notice that even common sense seems

FIGURE 4.8. U.S. Per Capita Copyrights

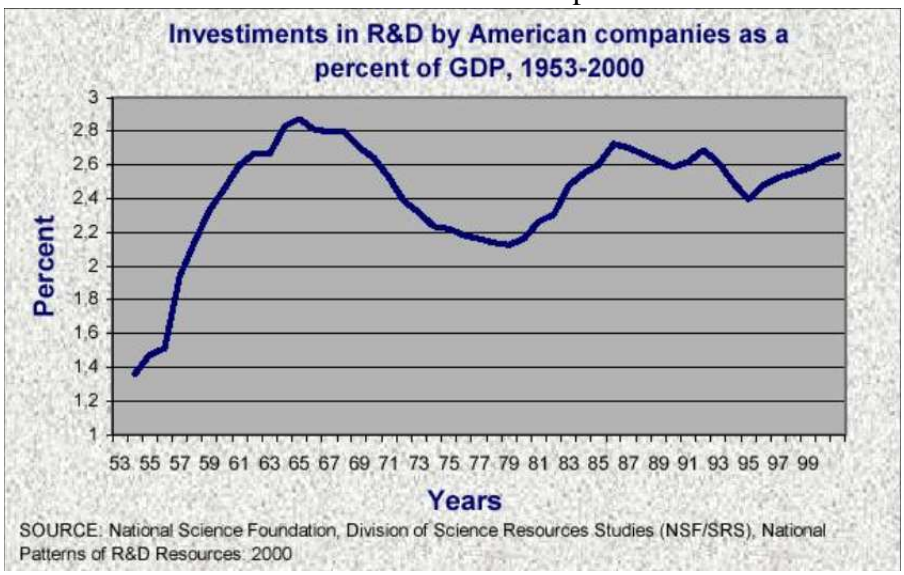


enough to reject the hypothesis that the elasticity of the total revenue function is constant or decreasing. In fact, this implies that whenever the market size doubles the share of population dedicated to creative activity more than double. Recall that, during the last century, the effective size of the world market has been multiplied, roughly, by a factor of one hundred. This implies that, by now, at least in the USA *everybody* should be working in the idea sector, which, alas, we have not yet achieved.

4.2.5. *Analysis of Labor Demand: Data.* With this background, let us look at some data. First we look at the time series. Here we must assume that the distribution $M(\rho)$ is not changing over time - for example, because all the good ideas have been used up. We also must assume that ϕ is either constant or increasing over time - as in fact it is. We look first at U.S. copyright data in Figure 4.8. We see that it grows significantly slower than U.S. population, and indeed is relatively static from the early 1900s to about 1970 - a period in which U.S. population more than doubles.

Turning to patentable ideas, R&D expenditure, insofar as it is an accurate measure and not simply a response to changing tax laws is a measure of labor in the idea sector. Figure 4.9 shows U.S. R&D expenditure as a fraction of GDP (the relevant size of market variable). Here from the early 1960s to the current time, this has remained relatively static - since U.S. has

FIGURE 4.9. U.S. R&D Expenditure



increased substantially over this time period, again this is strong evidence against a linear increase of labor in the idea sector.

Finally, we look at the cross-section. We examine R&D as a fraction of GDP across countries. If we assume that the domestic market is what is significant, we can combine data on GDP from the 1990 CIA World Factbook with R&D data and data on the strength of IPRs from Kanwar and Evanson [2001] (taking the average from the two periods) and regress the log of R&D as a fraction of GDP on the log λ. One issue that arises is whether we should measure λ by population or by GDP. Increases in per capital 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 capital income may mean twice as many ideas. We report results for GDP with the corresponding population coefficient in square brackets in Figure 4.10 - as it turns out it does not make much difference.

As can be seen, the elasticity is nowhere close to unity. However, this assumes 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

FIGURE 4.10. Cross Sectional Data

IPRs Strength	Size Elasticity [Population]	N	Average GDP USD billions	Exports
1	0.21 [0.21]	7	108	14
2	0.38 [0.22]	7	94	33
3	0.38 [-0.04]	12	394	85
4	0.13 [0.14]	4	1586	178

the same result as regressing on $\lambda_{domestic}/\lambda_{world}$, these 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 is on the order of 3-5, it is possible that the elasticity of per capital R&D with respect to size of market is near unitary. However, this ratio is implausibly large. Since world GDP is on the order of $20,586 billion, the potential size of the world market is much larger than 3-5 times the GDP of even the largest set of countries. However, this grossly overstates the relevant size of the world market. Exports are a fraction, not a multiple of GDP. Consequently a ratio of 3-5 would be possible only if R&D is much more intensive in export industries than domestic industries - by a factor of considerably more than 3-5. Using Lo's [2003] detailed data from Taiwan, in 1991 export intensive industries spent about 1.8 times as much on R&D 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 hardly supports the idea that $\lambda_{world}/\lambda_{average}$ is on the order of 3-5.

4.2.6. *Conclusion on the Elasticity of Total Monopoly Revenue.* We examine a variety of data from different sources, ranging from book revenues to patent values estimated by renewal rates, to R&D expenditures. We look both at cross-sectional and time-series data. Each individual analysis has many caveats - and this is clearly an area with a high rate of return to careful empirical work. However, given that all of the different sources of data say the same thing, we think it a fairly safe conclusion that in practice the elasticity of total monopoly revenue is increasing. Our best guess as to the functional form for $M(\rho)$ would be that it is approximately linear in the relevant range.

5. EXTENSIONS

5.1. **Return Nonneutrality.** In this section we drop the quality neutrality assumption and study the optimal degree of protection when private and

social values are not in a constant proportion to each other. Note that, as a function of private return under monopoly, social value under monopoly is $\overline{v}^M(\rho)\rho$ while social value under competition is $\overline{v}^C(\rho)\rho$ so that, for given ϕ, expected social value is $\rho[\phi\overline{v}^M(\rho)+(1-\phi)\overline{v}^C(\rho)]$. The two polar cases in which private return and social values are, respectively, positively and negatively related are worth considering.

If goods with lower private return have also lower social value, in the sense that $D\overline{v}^M(\rho) > 0$ and/or $D\overline{v}^C(\rho) > 0$, common sense and simple calculations show that this is a further reason for the length of protection to decline with the scale of the market. On the other hand, if either $D\overline{v}^M(\rho) < 0$ and/or $D\overline{v}^C(\rho) < 0$, or both, then this weakens the connection between the scale of the market and the declining optimal protection. In particular, in this case, it becomes possible to have the optimal degree of IP protection increasing with market scale, even when the elasticity of total monopoly revenue is increasing, so that the labor constraint never binds.

On the other hand, while $D\overline{v}^M(\rho) < 0$, $D\overline{v}^C(\rho) < 0$ may seem to reinforce the case for increasing IP protection, in fact it weakens it. That is, $D\overline{v}^M(\rho) < 0$, $D\overline{v}^C(\rho) < 0$ means that private return is poorly correlated with public benefit. In the extreme case, there may actually be a negative correlation between private and public benefit. In this case, the private sector produces the ideas of least social merit first, so it is certainly not worthy of IP protection when the scale of the market is small as this would simply lead to production of ideas of little social value. The argument for strengthening protection as the scale of the market increases, is that the increased scale of the market eventually leads the private sector to produce ideas that do have some significant social value, and at this point, we can try to compensate for the weakness of private incentives by increasing the level of protection. While this is formally correct, it clearly is a lopsided argument when it comes to designing welfare-improving policies. It takes as given the policy instrument, patents, even if the latter is the least adequate to maximize social welfare (because social value moves opposite to private value.) If it were really the case in practice that privately valuable innovations have little or no social value, and vice versa, then a form of government intervention other than IP would be recommendable, such as publicly sponsored research projects, or auctioning of production rights, or subsidies for innovators producing the socially valuable ideas, for example. Patents, certainly not.

Alternatives To Government Grants of Monopoly. The latter remarks, that when private and social values of new ideas are not aligned IP protection is the least appropriate policy instrument, suggest we should, albeit briefly,

consider how alternative forms of government intervention fare in our environment. An obvious alternative to having the government award private monopolies is to have the government award prizes for innovation. This can be financed in much the same way that private monopolies raise money - by imposing a sales tax on sales of new goods. Unlike the award of a private monopoly, the tax rate does not need to be set so high as to give the monopoly revenue, and Gilbert and Shapiro [1990] show in effect that having a low tax throughout the life ot the good is preferable to having a high tax (monopoly) for part of the life of a good and low tax (after the copyright/patent expires) for the remaining life of the goods. Hence, such a system of taxes, is intrinsically less distortionary than awarding private monopolies. Insofar as the prize money is simply paid back to the innovator, this is essentially the same as a system of mandatory licensing, in which the holder of the private monopoly is required to sell at a government mandated prices. Systems of mandatory licensing are widely used - in copyright, for example such things as radio play of music and xeroxing of copyrighted materials are covered by mandatory licensing provisions. In the case of patent, mandatory licensing was widely used in Taiwan until they were forced to reform their patent system by the United States. So this kind of mandatory licensing represents, as we might expect, the efficiency improvement from replacing an unregulated monopoly with a regulated monopoly.

However, there is little reason that the proceeds of taxes on new goods should be paid back to the innovator. From an efficiency perspective, it is better that the proceeds be used to defray the costs of producing innovations of high social value. This has several advantages over an intellectual property system. First, to minimize the monopoly/tax distortion, the minimum necessary to get innovation should be paid. In particular, it is best to pay $h(\omega)$, the indivisibility, to the innovator rather than the full social value. The intellectual property system makes little use of social knowledge of $h(\omega)$; with the exception of the non-obviousness requirement (now largely defunct) of patent law, patents and copyright base reward on social value rather than social cost. Second, as we noted above, if it is indeed the case that social value is poorly correlated with private value (the strongest case for increasing intellectual property protection as the scale of the market increases) a system of rewards based on other information about social value is likely to lead to a much better mix of innovations being produced. It is important to note that, like mandatory licensing, systems of public (and private) prizes have been widely used and are of demonstrated practicality. Historians of aviation have argued that prizes played an important role in the development of the airplane. The current X-prize has unleashed an enormous amount of innovation in aerospace technology, with the first privately funded supersonic flight taking place within the last year.

The issue, also in the context of our model, boils down to the public knowledge of the true social cost of introducing a new idea. When the latter is known, a public subsidy to innovators equal to the amount $h(\omega)$, financed by a consumption tax and followed by unconstrained competition, is easily shown to provide the least distortionary mechanism. When the information about the true cost $h(\omega)$ of innovating is private the problem appears less straightforward and worthy of further investigation.

5.2. **Consequences of Competitive Rents.** As argued in Boldrin and Levine [1999, 2002, 2004] it is by no means true that in the absence of any IP protection profits for innovators are negligible or even zero. At each moment of time, and especially shortly after innovation took place, a capacity constraint is present that will give rise to non-distortionary competitive rents. Most likely, there are also first-mover advantages, such as those documented by Tofuno [1989] in the market for financial securities.

We can model this by assuming a capacity constraint $\overline{z}(\omega)$ on per capita production after the IP protection expires. In this case, assuming the capacity constraint is not binding during the IP protection period, the innovator's total revenue from an idea ω is

$$\phi\lambda z^M(\omega)v_z(z^M(\omega),\omega)+(1-\phi)\lambda\overline{z}(\omega)v_z(\overline{z}(\omega),\omega).$$

Assuming the labor constraint does not bind, hence $w=1$, dividing through by $h(\omega)$, we may write the condition for good ω to be produced as

$$\phi\lambda\rho(\omega)+(1-\phi)\lambda\rho^C(\omega)\geq 1$$

where $\rho^C(\omega)=\overline{z}(\omega)v_z(\overline{z}(\omega),\omega)/h(\omega)$ is the competitive rent per unit of indivisibility cost. Let us use the simplifying assumption that competitive rent is proportional to monopoly revenue, per unit of indivisibility cost; that is $\rho^C(\omega)=\vartheta\rho(\omega)$, with $0<\vartheta<1$. Note that this is stronger than our earlier neutrality assumptions, which had to hold only in expected value. Then we may again write social welfare entirely in terms of ρ and the only modification of our earlier expression for social welfare is that $\underline{\rho}=[(\vartheta+\phi(1-\vartheta))\lambda]^{-1}$. The corresponding NOC is

$$\begin{aligned}NOC(\lambda,\phi)=&(1-\vartheta)\left[(1/(\vartheta+\phi(1-\vartheta)))\left\{\phi\overline{\mathrm{v}}^M+(1-\phi)\overline{\mathrm{v}}^C\right\}-1\right]\\&\times\Upsilon(1/(\vartheta+\phi(1-\vartheta))\lambda)\\&-(\overline{\mathrm{v}}^C-\overline{\mathrm{v}}^M)\end{aligned}$$

from which we see that it continues to be true that if

$$\Upsilon'(1/(\vartheta+\phi(1-\vartheta))\lambda)>0$$

then $\hat{\phi}(\lambda)$ is unique and non-decreasing. In this case, we may also check that $NOC_{\vartheta} < 0$, so that higher competitive rents lead to a reduction in the optimal level of protection.

Finally, consider that the NOC at $\phi = 0$ is

$$(1-\vartheta)\left[(1/\vartheta)\overline{v}^{C} - 1\right]\Upsilon(1/\vartheta\lambda) - (\overline{v}^{C} - \overline{v}^{M})$$

and as $\lambda \to \infty$ this approaches $-(\overline{v}^{C} - \overline{v}^{M}) < 0$. Hence, when competitive rents are present, the optimal level of protection should be set equal to zero at a finite market size, not just asymptotically as in the extreme case of no competitive rents.

5.3. **Consequences of Rent-Seeking.** Suppose that the size of the indivisibility does not vary systematically with private return. Then ideas with high returns also have high absolute levels of profit associated with them. Suppose it is possible to purchase "extensions" of protection, in the form of a $\Delta\phi$, from the government sector at a cost. Then it is owners of ideas with high ρ that have the greatest incentive to do so, as they can "leverage" the $\Delta\phi$ more than anyone else. This means that the marginal firms, who from a social point of view are the reason for IP protection , do not get much say over the length of protection. In the extreme case the marginal firms get no protection, so the set of ideas produced is the same as without IP, and IP serves only to introduce a monopoly distortion. Rather remarkably, Lerner and Posner [2001] suggest embodying such a scheme in law.

5.4. **Variations on the Utility Function.** We assumed that utility is linear in the output of the idea sector and in labor. We can consider more generally, the functional form

$$U\left(g(\lambda)\int_{\underline{\rho}}^{\infty}[\phi\overline{v}^{M} + (1-\phi)\overline{v}^{C}]\mu(\rho)d\rho\right)$$
$$+V\left(L-(1/\lambda)\int_{\underline{\rho}}^{\infty}\mu(\rho)d\rho\right).$$

The empirically relevant case is that in which $U(\cdot)$ displays, after a point, sharply decreasing marginal utility while $V(\cdot)$ has displays a fairly high marginal utility even at L. To see how this works, examine first the sharp case in which U is linearly increasing to an upper limit, then constant, while V is linear all thorough its domain. In other words, once total output in the idea sector is large enough, no further utility is generated, or $U(x) = x$ for $x \leq X$ and $U(x) = X$ for $x > X$. This is a simple way of modeling the intuitive case, following from the assumption that $g(\lambda)/\lambda$ is increasing in

λ, in which marginal utility of consumption decreases faster than the marginal utility of leisure as the former goes to infinity and the latter remains bounded above by L. As long as λ is small enough, satiation does not affect the original equilibrium, and the price of output in the idea sector is $P = 1$. However, once λ grows large enough, the satiation constraint binds.[5] Equilibrium requires that exactly X be produced in the idea sector, so the price in the idea sector $P < 1$ must fall to discourage labor from flowing into that sector. But a more efficient way to discourage labor from flowing into the idea sector is simply to lower ϕ. This simple result is the mirror image of the labor constraint binding in the previous analysis.

More general non-linearities in U have a similar effect as long as our basic assumption about relative marginal utilities of consumption and leisure holds. The latter can be reformulated in the following way: as the size of the market grows and more ideas are produced, the price in the idea sector declines faster than the marginal utility of leisure for fixed ϕ, hence the relative price of labor goes up. This seems to be reflected, in the real world, by the century-long decline in hours worked per capita. Notice that it is immaterial to our argument if the increase in the size of the market is due purely to an increase in the population, or to an increase in trade, or to unmodeled technological advances that reduce the size of the indivisibility $h(\omega)$. In general it will be best to exploit the opportunity offered by an increase in the size of the market by reducing ϕ, rather than by allowing the relative price of labor to rise as the latter does not engender any efficiency gain but purely redistributes wealth from the general population to the skilled labor used in the production of ideas. The effect of V being non-linear is similar. As leisure declines, wages of skilled workers rise gradually relative to the wage of people not employed in the idea sector. Like before, it is best to offset this purely re-distributional effect by reducing ϕ.

5.5. **Positive Marginal Cost of Distributing Ideas.** So far we have assumed that there is no marginal cost of reproducing and distributing ideas or goods that use ideas (for short: reproduce ideas.) There are several possible cases, depending on which inputs are needed to do this. One possibility is that the same labor used to create ideas is used to reproduce them. This case is rather complex, because it introduces a third margin into the choice of ϕ - the monopoly for inframarginal ideas, the marginal ideas, and the amount of labor used to reproduce existing ideas.

However, in practice the type of labor used to reproduce ideas is probably not a terribly good substitute for the labor used to produce the ideas themselves. If we introduce an additional factor of production - unskilled

[5]Notice that $U'(X) < V'(L)$ is pretty much all that is needed here.

labor, call it - and assume that this is used to reproduce ideas, provided that this factor is in plentiful supply so that marginal cost is constant, little change is needed in our analysis. In particular, instead of examining $v(z,\omega)$, we should examine per capital utility net of the cost of reproducing the idea: $v(z,\omega) - mc(\omega)z$. This may have an impact on whether quality is neutral, since it may be neutral for $v(z,\omega)$, but not for $v(z,\omega) - mc(\omega)z$, but, for example, with quadratic utility (linear demand) we have neutrality in both cases. In general, we would expect this modification to make it more likely that private return and social value of an idea move together. This is because one intuitive reason why social value increases with private return is that the marginal cost of producing additional copies of an idea increase more slowly than utility from the additional copies. As we have seen in Section 5, this reinforces our main prediction, that IP protection should decrease when market size increases.

If the unskilled labor constraint binds before the skilled labor constraint does - something we think is unlikely, but since the demand for unskilled labor in this model grows much more rapidly than the demand for skilled labor as λ increases, something we recognize is a possibility - then the situation again is complicated by the existence of a third margin. Unskilled labor must be allocated between producing more copies of particular goods, versus producing more kinds of goods. The point is, though, that the central finding from our analysis of the basic case does not change. As the size of the market increases, if the unskilled labor constraint binds before the one for skilled labor does, the choice is between shifting more unskilled labor to activities that make use of new ideas or to activities that make use of old ideas. When the marginal social value of the latter is larger than that of the former a reduction in the level of IP protection ϕ is recommended.[6] When the opposite is true, and assuming again that private and social values are properly aligned, unskilled labor becomes, in this model, a re-scaled version of skilled labor and the previous analysis, which recommends a reduction in ϕ when the size of the market increases, applies.

6. International Trade and Harmonization

We now turn to the issue of IP protection in the world economy. We assume that there are k countries and that each country i has a fixed fraction θ_i of world demand, labor and world ideas. The total size of the world economy is still λ.We examine the case in which countries may not discriminate against foreign inventors. While *de-facto* violated in some occasions, this reflects current legal practices around the world, and it allows us to focus

[6]Also in this case, as before, should the private and social values move in opposite direction the analysis acquire a further degree of complication.

on the specific role of IP protection. We let ϕ_i denote the level of IP protection in country i. Throughout this section we assume there is complete and costless free trade of goods, that the labor constraint does not bind, and that the elasticity of total monopoly revenue is increasing.

From an inventor's perspective, what is relevant is the effective (weighted by market shares) total protection received world wide. This is simply $\phi = \sum_i \phi_i \theta_i$, hence $\underline{\rho} = 1/\phi\lambda$ determines the marginal invention worldwide. Each country is assumed to maximize own social welfare, which is given by

$$g(\lambda) \int_{\underline{\rho}}^{\infty} [\phi_i \overline{\nu}^M \rho + (1-\phi_i)\overline{\nu}^C \rho - 1/\lambda] \mu(\rho) d\rho + L,$$

the significant feature of which is that the monopoly distortion depends on ϕ_i and not on ϕ. Note that in this world, each country is assumed to get a *pro rata* share θ_i of the total mass of world innovations, so that net revenue to a country from selling/purchasing IP-protected products on the world market is zero.

With standard notation, the NOC corresponding to the social welfare problem for country i is

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

Because the elasticity of total monopoly revenue is assumed increasing, this is strictly concave in ϕ_i and continuous as a function of (ϕ_i, ϕ_{-i}) so has a pure strategy Nash equilibrium, characterized by the first order conditions $NOC(\phi_i, \phi_{-i}) = 0$.

Consider first a symmetric equilibrium of a symmetric model in which $\theta_i = 1/k$. In equilibrium we must have $\phi = \phi_i$, or

$$NOC_k(\phi) = (1/k) \left[(1/\phi) \left\{ \phi \overline{\nu}^M + (1-\phi)\overline{\nu}^C \right\} - 1 \right] \Upsilon(1/\phi\lambda) - (\overline{\nu}^C - \overline{\nu}^M) = 0.$$

Holding constant the total market size λ, let ϕ^1 be the solution to the single country problem and $\phi^k$the symmetric solution to the symmetric k country problem. Because the first term in $NOC_k(\phi)$ is positive and $\theta_i < 1$ we have that $NOC_k(\phi^1) < 0$, implying, since $NOC_{k\phi} < 0$ under the elasticity condition, that $\phi^k < \phi^1$. Notice that ϕ^k is decreasing in k so that, as the number of countries increases the symmetric Nash equilibrium converges to the case of no IP, which is suboptimal in our setting. The intuition behind this result is simple: by decreasing ϕ^k a country loses because it creates fewer new goods and gains because it consumes at the competitive level the goods created by the remaining $k-1$ countries. As k increases the second margin strictly dominates the first.

The concavity of the objective function in ϕ_i implies that each country choosing the solution to the single country problem is in fact at the unique

social optimum for the world as a whole. Moreover, if each country is constrained to set the same level of protection as all others, for example through a legal mechanism such as the WTO, they would all agree to choose the social optimum ϕ^1. This is the standard harmonization result: in the unconstrained protection game countries under-protect due to the public goods nature of IP protection, and a WTO-like mechanism that forces harmonization leads them to the second best.

Unfortunately, this analysis, while mathematically correct, has little relevance to policy analysis for two reason. First, it assumes that countries choose policies to maximize their own social welfare. The evidence of vastly increased levels of protection in response to vastly increased market size provides relatively conclusive evidence that this is not how countries choose their levels of protection; rather it suggests that levels of protection are determined largely by rent-seeking. The fact that copyright extensions have applied retroactively is yet another piece of evidence that the level of protection is driven by rent-seeking and not social efficiency. Second, current extensions of IP are not between countries of equal size with currently equal levels of IP, who by harmonizing will agree to a common higher and more efficient level of protection. Rather, extension of IP protection is taking place between large countries with high levels of protection and small countries with no protection. To study this, the relevant model is not one of k identical countries, but rather of one large country (the U.S. and the E.U.) and $k-1$ identical small countries.

Consider then, a situation where there is one large country with share θ_1 and $k-1$ small countries with shares $\theta_i = (1-\theta_1)/(k-1) < \theta_1$. First we show that, in this case, regardless of k the equilibrium level of aggregate protection, ϕ, is bounded away from zero. The NOC for the large country is

$$NOC_1(\phi_1,\phi) = \theta_1 \left[(1/\phi)\left\{\phi_1\overline{v}^M + (1-\phi_1)\overline{v}^C\right\} - 1\right] \Upsilon(1/\phi\lambda) - \overline{v}^C - \overline{v}^M) = 0.$$

Observe that $\phi \geq \theta_1\phi_1$ and recall that $NOC_1(\phi_1,\phi)$ is decreasing in ϕ. Hence, $NOC(\phi_1,\theta_1\phi_1) \geq 0$. Since this latter expression is also decreasing in ϕ_1 a solution to $NOC(\widetilde{\phi}_1,\theta_1\widetilde{\phi}_1) = 0$ must satisfy $\phi_1 \geq \widetilde{\phi}_1 > 0$. This in turn implies that in equilibrium $\phi \geq \theta_1\widetilde{\phi}_1 > 0$. This shows that ϕ is bounded away from zero independent of k because the large country will never impose a negligible amount of protection.

We now turn to the NOC for the small countries. At $\phi_i = 0$ this is

$$\begin{aligned} NOC_k(0,\phi) &= (1-\theta_1)/(k-1)\left[(1/\phi)\overline{v}^C - 1\right]\Upsilon(1/\phi\lambda) - (\overline{v}^C - \overline{v}^M) \\ &\leq (1-\theta_1)/(k-1)\left[(1/\theta_1\widetilde{\phi}_1)\overline{v}^C - 1\right]\Upsilon(1/\theta_1\widetilde{\phi}_1\lambda) - (\overline{v}^C - \overline{v}^M) \end{aligned}$$

which is strictly negative for k larger than

$$K = \frac{(\overline{\mathrm{v}}^C - \overline{\mathrm{v}}^M)}{(1-\theta_1)\left[(1/\theta_1\widetilde{\phi}_1)\overline{\mathrm{v}}^C - 1\right]\Upsilon(1/\theta_1\widetilde{\phi}_1\lambda)} + 1.$$

Since there is always a unique solution to $NOC_k(\phi_i, \phi) = 0$, for $k > K$ it occurs at $\phi_i = 0$. In this case the solution for the large country is simply the solution to the social optimum problem which ignores demand and supply from the rest of the world; that is, the choice of ϕ_1 is the one that would be optimal for a population of $\theta_1\lambda$. By the usual scale of market effect, that means the equilibrium solution for $\phi_1$is larger than the value that maximizes world social welfare, that is, the solution to the social optimum problem with population λ.

In short, in the empirically relevant case where there are small countries with no IP and large countries with excessive IP, effective harmonization requires not only that all countries be constrained to set the same level of IP protection, but that the level of IP protection chosen be substantially less than the existing levels of protection in the large countries.

Notice, incidentally, that our base assumption is that countries cannot increase the level of domestic innovation by changing their IP laws, as their share of total innovation is fixed at θ_i. Insofar as countries can increase their share of world-wide innovative production by changing their national level of IP protection, they benefit from the fact that they get a disproportionate share of the total revenue from innovation. In fact there is some evidence that favorable IP treatment can attract innovation. There are several reasons for this. First, favorable IP legislation may be a signal of favorable treatment of innovators in general (for example, as in Ireland, through tax law). Second, although legal discrimination against foreign inventors is forbidden in principle, there may be a variety of informal reasons why it is advantageous to be a domestic innovator to take advantage of strong local IP protection. Finally, the distribution of innovation across countries can be driven by the explicit rent-seeking behavior of innovators, who may choose to reward countries that provide favorable IP protection with increased revenue from domestic innovation. The movie industry, for example, has gone to some length to reward and punish its political foes with movie production.

Insofar as increasing IP protection lures innovation, a second type of equilibrium distortion arises. Rather than underprotecting in an effort to free ride off of innovation in other countries, the incentive is to overprotect to try to get a disproportionate share of IP revenue.

7. Conclusion

7.1. **Relation with Previous Literature.** Besides the work of Grossman and Lai [2002, 2004] there is a wide partial equilibrium literature on the optimal length of patent protection. This literature, stemming from the paper of Gilbert and Shapiro [1990] examines the trade-off between patent length and breadth for a single innovation. Gilbert and Shapiro give assumptions under which optimal length is infinite, while Gallini [1992] shows that with a more realistic model of the "breadth" of protection, this result may be reversed. This literature does not examine the broader question of optimal policy that covers many different ideas, and takes as given that policy may easily determine "breadth" as well as "length." We think that "breadth" is much more difficult to legislate than "length," and because it is less visible, more subject to rent-seeking, regardless of legislative intent. In our model, unlike this literature, we effectively take the "breadth" of protection as exogenous and focus on length. Insofar as "breadth" as well as "length" can be legislated, our parameter ϕ may be regarded as kind of a summary of length and breadth combined. Hence, it would be good public policy to reduce breadth as the scale of the market increases as well as length. In particular, it might be a good idea to introduce the "independent invention" defense as suggested by Scotchmer [2002], or to eliminate product patents in favor of process patents only. However, it is clearly more practical to tie a time limit to the growth of the economy than a particular scope of coverage.

A number of reasons, other than the invariance of the optimal length of patents to the size of the market, have been invoked to justify policies asking for an increase of the degree of IP protection as a consequence/condition of trade liberalization. First, as Diwan and Rodrik (1991) argue, northern and southern countries generally have different technology needs and, without the southern protection of IPRs, northern countries would not develop technologies largely needed by the South. This seems rather at odds with both basic principles and historical facts. Developing countries have developed, almost always and invariably, by adopting technologies that had been developed by the advanced ones. Indeed, the very same notion of "convergence" would make no sense if this were not the case. Examples of countries that have developed late by inventing technologies substantially different from those invented by the already developed ones, are conspicuously absent. China and India are currently achieving unusually high growth rates by adopting *exactly* the same technologies adopted earlier by developed countries; this is currently called "outsourcing," and it means just that. So much for history and facts. But even in an abstract and purely theoretical sense, the Diwan and Rodrik story can make sense if and only if one argues that poor countries have natural and un-modifiable resources

and skills that are completely different and orthogonal to those that richer countries had when they were at the same stage a development. Further, one needs to also assume, or show, that the technologies adopted by advanced countries can neither be learned, or transported to the poorer ones, or that doing so is clearly more costly than developing completely new, and as of now yet unknown, technologies. Both claims sound rather implausible. All historical evidence, including the development of the USA in the nineteenth century, shows that "convergence" or "catching-up" takes place only when the followers imitate the technologies of the leaders. Hence, facilitating technological imitation (call it "pirating" if you like) is *the key* to economic development for newcomers. No imitation, no (development) party.

A second line of thought argues that northern firms may react to the lack of IPRs in the South by making their technologies more difficult to imitate, which can result in less efficient research technologies and less northern innovations (Taylor, 1993, 1994; Yang and Maskus, 2001). This is possible, in principle; again facts seem to contradict it, though. In theory, the very opposite effect is also possible: the more the South imitates the North, the more the Northern firms will react by innovating in order to stay ahead and preserve the value of their fixed or immobile factors. In this case, lack of IP protection in the South fosters higher rates of innovation in the North. Indeed, this seems exactly what is going on and has always gone on in the history of economic development. Like it or not, imitation is the key ingredient of competition, a fact economists have recently tried to forget, but which was quite clear to Adam Smith. Historically, IP protection in the underdeveloped countries was even weaker than now in the nineteenth century, and this does not seem to have lead firms in advanced countries to invest heavily in secrecy and/or innovate less. If secrecy was so high when, say, the U.K. was the leading country and most of economically backward Europe had no IP protection laws, how did those European countries ever manage to imitate and adopt the British technologies? The same, rhetorical, question could be asked, for Japan, South Korea and a long list of now developed countries.

Third, some commentators on intellectual property, such as Landes and Posner [2003] have become confused on the point of optimal length of intellectual property protection, arguing for a system of perpetual copyright renewal that would have little effect in increasing the incentive to innovation, while perpetuating the monopoly losses on inframarginal productions. There is no sense in which such informal arguments, often developed by confusing "pecuniary" with "real" externalities, can be derived from a general equilibrium theory of optimal IP protection.

Finally, it should be noted that the theory proposed here does not say that the South should have systematically less IP protection than the North but,

instead, that both should have a common IP protection at a level substantially lower than it is currently in the North, e.g. the USA and the EU.

7.2. **Shortcomings and Future Research.** The most important missing aspect of our analysis is the dynamic feature that ideas build on other ideas. As pointed out in Scotchmer [1991] and Boldrin and Levine [1999, 2003], ideas that use other ideas as inputs greatly weaken the case for IP because the latter, while it encourages innovations by improving the return to the first inventor, discourages further innovations through raising their cost. In this sense, there is no reason to think that adding true dynamic features to the model is likely to make IP more socially desirable. In fact, when the complexity of innovations increases because new ones need to use more and more old ideas as inputs, the presence of widespread IP naturally determines a hold-up problem where even one residual monopolist may prevent new ideas from being implemented (see Boldrin and Levine [2004b] for a simple formal version of this argument.)

As we observed, neither that the number of ideas may increase with size at different rates for different characteristics, nor that the indivisibility varies with the size of the economy are possibilities considered here. These are valuable, but theoretical demanding, extensions for various reasons. The first extension is valuable because, to the extent that new ideas are not all equally useful, the fact that market size may make ideas with certain characteristics more abundant than other is relevant for welfare. The relevance comes from the fact that ideas with high private returns need not be ideas with high social values, and viceversa; an optimal innovation policy should be designed by taking this effect into account. This is a hard problem to tackle on purely theoretical grounds, and, unfortunately, we are aware of no empirical research measuring the relationship between private return and social value of new ideas.

The second extension, allowing for variations in the indivisibility $h(\omega)$ that are due to variations in market size, is also interesting. In theory, one may argue that a larger economy implies more competition to develop ideas potentially already available, and few additional potential ideas; that is $h(\omega, \lambda)$ increases with λ while $g(\lambda)$ is practically flat. The opposite case is, clearly, also a theoretical possibility. Which of the two is empirically relevant is important to provide guidance to theoretical research. To discriminate between the two polar hypotheses one needs to be able to measure the portion of $h(\omega)$ that is wasted when one does not win the innovation race, versus the complementary portion that may be used to come up with further and different innovations. That is, how much do past, even unsuccessful, research efforts contribute to future successes? Anecdotal evidence is mixed, and serious empirical work is completely missing. Similarly, the function

$g(\lambda)$ also needs to be measured. Are we all drawing from the same Platonic urn of ideas? Does schooling allow us, at least, to sample without replacement, so that more people means faster drawing of new ideas? Or are we, instead, sampling from the same urn with replacement, in which case more and more useless duplicates are drawn as λ increases? Or, maybe, adding more people increases the number of original ideas in the urn from which we sample, making $g(\lambda)$ increase quite fast? These are important empirical and theoretical questions that current research has not yet considered.

Turning to the policy implications of our results: if $\phi\lambda$ is held fixed, the quality of ideas produced remains unchanged. This is not the social optimum: generally we will want to take advantage of the increased λ to allow some marginal ideas to enter the market. A simple rule of thumb is to observe that for large λ a linear $M(\rho)$ implies that protection should be reduced by about half the increase in scale of market. Thus, the simple rule of thumb would be that if the size of market doubles, the amount of protection should be reduced by about 1/3 - that is, a 50% increase in 2/3rds the level of protection is half of the 100% increase in the size of the market. Taking a real interest rate of 2% per year, observe that $1-\phi = e^{-rT}$ where T is the length of IP protection. Using $\log(1+x) \approx x$ this gives $\Delta T \approx 50\Delta\phi$.

For example, the G7 nations account for about 2/3rds of world GDP. If we think of the intellectual property changes in the WTO as extending the protection that exists in the G7 to the rest of the world, this suggests a reduction in protection by about 1/6th. This means approximately an 8 year reduction in term. Similarly, if the world economy is growing at 2% a year, a simple rule of thumb would be to reduce protection terms by 1% per year, or about 6 months per 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 millions copies. The effective size of the market has, therefore, increased of 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 of about a factor of twenty-five. 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 of grown of 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 about 6 months, rather than the current term of approximately 100 years. This gives a ratio of two hundred between the actual copyright terms and their socially optimal value.

Our results are relevant for the debate on the impact of IP harmonization policies on developing countries. Romer [1994], among other, has pointed out that in the presence of fixed costs the welfare loss from tariff protection may be orders of magnitude larger than the usual Harberger's triangle. The same logic applies to the welfare losses due to IP protection. The point to notice here, is that monopoly prices due to patents and copyright have the same effect as taxes on ex-post profits and lead to the non-creation of a new good or the non-adoption of a new production process when the indivisibility is relevant.

REFERENCES

[1] Acemoglu, A. and F. Zilibotti (1996), "Was Prometheus Unbound by Chance? Risk, Diversification and Growth, *Journal of Political Economy,* **105**, 709-751.

[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*,

[5] Boldrin, M. and D.K. Levine (2004a), "IER Lawrence Klein Lecture: The Case Against Intellectual Monopoly,*The International Economic Review,* vol (May).

[6] Boldrin, M. and D.K. Levine (2004b), "The Economics of Ideas and Intellectual Property," mimeo, University of Minnesota, UCLA, and Federal Reserve Bank of Minneapolis, October.

[7] DiMasi, J. A., R. W. Hansen, H. G. Grabowski, L. Lasagna (1991), "The Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics,* **10,** 107-142.

[8] Diwan, I. and D. Rodrick (1991), "Patents, Appropriate Technology, and North-South Trade," *Journal of International Economics,* Vol.30, pp.27-48.

[9] Gallini, N. (1992), "Patent Policy and Costly Imitation," *Rand Journal*, **23,** 52-63.

[10] Gilbert, R. and C. Shapiro (1990), "Optimal Patent Length and Breadth," *Rand Journal*, **21,** 106-112.

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

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

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

[14] Grossman, G. M. and E. L. Lai (2002), "International Protection of Intellectual Property," NBER.

[15] Grossman, G. M. and E. L. Lai (2004), "International Protection of Intellectual Property," mimeo Princeton.

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

[17] Hart, O. D. (1979), "On Shareholder Unanimity in Large Stock Market Economies," *Econometrica,* **47,** 1057-83.

[18] 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.

[19] Kanwar, S. and R. E. Evanson (2001), "Does Intellectual Property Protection Spur Technological Change?" Cowles.

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

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

[22] 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.

[23] 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.

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

[25] Makowski, L. (1980), "Perfect Competition, the Profit Criterion and the Organization of Economic Activity," *Journal of Economic Theory,* **22**, 222-42.

[26] Maurer, S. M. and S. Scotchmer (2002), "The Independent Invention Defense in Intellectual Property,*Economica,* forthcoming.

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

[28] Romer, P. (1994), "New Goods, Old Theories, and the Welfare Costs of Trade Restrictions," *Journal of Development Economics*, **43**, 5-38.

[29] Sampat, B.N., and A.A. Ziedonis (2002), "Cite Seeing: Patent Citations and the Economic Value of Patents," mimeo, Georgia Institute of Technology and University of Michigan, november.

[30] Scotchmer, S. (1991), "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives,* **5.**

[31] 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.

[32] Taylor, M. S. (1993), "TRIPS, trade, and technology transfer," *Canadian Journal of Economics,* Vol.26, No.3, pp.625-637.

[33] Taylor, M. S. (1994), "TRIPS, trade, and growth," *International Economic Review,* Vol.35, No.2, pp.361-381.

[34] Yang, G. and M., Keith E. (2001), "Intellectual Property Rights, Licensing, and Innovation in an Endogenous Product Cycle Model," *Journal of International Economics,* Vol.53, pp.169-187

APPENDIX: BOOK REVENUE DATA

We collected all the titles, ISBN numbers, and sale prices listed by www.amazon.com for the query *hardcover fiction books* and for the two publication periods of September 2003 and September 2004. The sales data is 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 sale 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 1/6 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.