Competition and Innovation

Michele Boldrin
Juan Correa Allamand
David K. Levine
Carmine Ornaghi

ABSTRACT

Which kind of intellectual property regime is more favorable to innovation: one that enforces patents or one that does not? Economic theory is unable to answer this question, as valid arguments can be made both for and against patents; hence we must turn to empirical evidence. In this paper, we review empirical evidence gathered by other researchers and add new evidence of our own. We conclude that the evidence suggests that patents do not promote innovation, but instead retard it.

Michele Boldrin is the Joseph Gibson Hoyt Distinguished Professor of Economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. Juan Correa Allamand is a researcher at the Ministry of Finance, Chile. David K. Levine is the John H. Biggs Distinguished Professor of Economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. Carmine Ornaghi is a lecturer in economics at the University of Southampton.

Cato Papers on Public Policy, Vol. 1

Competition and Innovation

1. INTRODUCTION

Our concern is with a straightforward policy question: Which kind of intellectual property regime is more favorable to innovation, hence more conducive to technological progress, increasing factor productivity, and—eventually—economic growth? More precisely, what is the role that patents play, or do not play, in fostering technological innovation? A practical policy question such as this rarely has a straightforward answer; this may be one of those rare cases when it does. Our final answer, in fact, will be that there is no evidence whatsoever that patents and copyright foster innovation and creation at large, while there is abundant evidence that they damage it in particular instances, apart from inducing substantial efficiency, rent-seeking, and distributional costs.

To allow for a focused discussion, we will pretend that a starker difference between legal regimes exists (and is the object of contention) than is probably feasible in reality, where legislation must act on a multidimensional continuum of choices. On the one hand, we consider a regime in which patents are assigned easily and then strictly and forcefully enforced for a long time—as has been the case in the United States for more than a century and increasingly so since the middle 1980s. On the other hand, we consider a counterfactual regime in which patents are hard to obtain, loosely enforced, and last for brief periods of time only, that is, a system in which imitation, and the free-entry competition it brings about, rules. Subject to the qualifications discussed in the paper, we call the latter arrangement ''competition'' and the former one ''monopoly.'' We ask: Which regime is more desirable from a social point of view?

At a theoretical level, this question receives two straightforward—but completely opposite—answers. Within a certain class of models, ''monopoly'' is the necessary evil without which innovation is impossible. Within a second class of models, ''competition'' nicely provides innovation and, in some cases, even reaches first-best.

Depending on the kind of assumptions one is willing to make about the set of available technologies, the cost of imitation, and the speed at which productive capacity can be built, economic theory may go either way. In one case—when innovations are dramatic and well identifiable, the cost of technological discontinuities is very high, productive capacity is easily built, and imitation is costless—intellectual property is likely to be a growth-enhancing arrangement. This is recognized today as the ''received wisdom.'' In the opposite case—when innovations are a distributed and incremental phenomenon the cost of which, while sometimes high in the aggregate, is relatively low at each single step, productive capacity cannot be rapidly adjusted, and imitation is costly—patent protection is likely to be a growth-reducing arrangement. This latter view is considered, nowadays, a form of ''economic heresy.'' It happens to be a view of the innovation process that, among others, two of us have been suggesting as appropriate for quite a while (see, e.g., Boldrin and Levine 2008b).

The question of which regime actually induces more innovation and is more socially desirable has been hardly considered at the empirical level, where most researchers have been taking it for granted that patents and copyright are the best incentives to innovation, hence to economic growth. Because, at the end of the theoretical day, the issue is an empirical one, it is to the small amount of available empirical evidence—to which we add a bit of fresh material—that we dedicate the bulk of this paper. Our examination leans strongly toward the heretic option: all things considered, patents do not have a favorable impact on technological innovation and tend to hurt economic welfare. Nevertheless, we will not even try to claim the issue can be settled on the basis of the available evidence, if anything because the latter is still scarce and unsystematically organized. As mentioned, maybe because so little relevant data have been collected or maybe because the received wisdom has been so dominant for so long that empirical verification appeared redundant, the number of empirical papers addressing the policy question we have asked turns out to be very small. Better data and more systematic analysis are therefore needed before any conclusive policy prescription may be drawn.

We proceed through the following steps: In Section 2, we start with a brief summary of the main theoretical features of the received wisdom. That wisdom has two sharp and testable predictions:

(a) The stronger the monopoly power granted by patents, the higher the incentive to innovate, hence the faster the growth rate of productivity.
(b) The innovation process consists of a sequence of discrete jumps; at each step a new innovator ''jumps over'' the incumbent monopolist, acquires a patent to protect her new innovation, and becomes a temporary monopolist, soon to be overtaken by an even better newcomer.

The empirical support for these predictions is considered in Section 3, and we find it to be scant. In fact, we show that neither (a) nor (b) receives any support in the statistical or historical literature. More cogently, empirical research has focused on two lines of investigation:

- The ''patent puzzle''—that is, the fact that, in recent decades, as patent protection was progressively strengthened, we did not observe an innovation explosion.
- The ''inverted-U curve''—that is, the claim that, in the data, while it is apparent that strong monopoly power does not favor innovation, free competition may not be all that good either, and innovation is maximized somewhere in the middle.

While we have little to add to the interesting findings collected in the patent-puzzle literature, we believe that something more can be said about the inverted-U claim. Here we make two points, one general and the other specific. The general point is that using the number of patents (or their citations count) to measure the actual number of innovations (or, better, their impact on productivity) across sectors and firms may lead to somewhat biased results. While many innovations lead to a patent, at least equally many others do not; while many patents are associated with an actual innovation, at least equally many others are obtained for legal or rent-seeking purposes that have little or nothing to do with actual innovations. As we show, while patent citations may be vaguely correlated with productivity growth, they are far from being a reliable measure of it. This suggests treating with the greatest care those studies that assume a close identity between the two: showing that, in the data, X may increase patents is a long shot from showing that X also increases innovations and labor productivity. The specific point—developed in great detail in the Ph.D. dissertation of one of us

(Correa 2010)—is that the inverted-U finding does not actually withstand a careful statistical examination and disappears rather quickly. When we reexamine the very same dataset in which the inverted-U relation was claimed to exist, we find instead a robust monotone relation between measures of competition and the number of patent citations. As the heretic view predicts, this relationship is increasing.

Section 4 summarizes the heretic line of thought according to which imitation does not hamper, and in fact stimulates, innovation. The heretic view predicts that we should observe a higher rate of innovation in those sectors in which competition is stronger and patent protection weaker. Because, as we argue, the heretic line of thought rests on the assumption that most technological progress is embodied in either human or physical capital, we also briefly discuss the cumulated empirical evidence supporting this claim.

In Section 5, we test our theory using recent high-quality data. By merging three different micro datasets on firms' patents, patent citations, and productivity growth at the sectorial level, we show there is a clear, monotone, increasing relationship between measures of sectorial competition, the innovative activity of firms, and—most importantly—labor productivity and total factor productivity (TFP) growth. The implications of our findings for future research and actual policymaking are collected at the end, in Section 6.

2. RECEIVED WISDOM

The intellectual roots of the currently dominant consensus about the role that patents and monopoly power play in the innovation process were planted long ago: Schumpeter (1942) ought to be seen as its starting point, while Scherer (1990) is a leading current rendition among industrial organization scholars.

A famous paper by Kenneth Arrow (1962) provided debatable but authoritative information-theoretical support for the view that ''innovation is a public good.'' That view, in turn, gave rise to two completely opposite lines of research and related policy implications. According to one line of thought, if innovation is a public good, then it cannot be provided by the markets, be they competitive or monopolized; innovation needs, therefore, to be either produced directly by government agencies or, at least, subsidized by means of fiscal revenues and tax incentives. This was the line of thought Arrow's initial argument was meant to support and that still appears as a logical consequence of the (in our view patently falsified) view that innovation is a public good. According to the second line of research stemming from Arrow's assumption, innovation becomes an undersupplied public good because the property right system is poorly designed: those contracts through which innovation can be made excludable are not available in a competitive market system as long as the latter allows for imitation. Patents can solve this particular version of the "tragedy of the commons": they make innovations excludable, thereby insuring at least a second-best outcome—Gilbert and Shapiro (1990) contains a relatively recent, and more elaborated, exposition of this point of view. This approach, which had come to dominate theoretical and applied work in the area of industrial organization by the late 1960s, was incorporated into dynamic models of economic growth and capital accumulation with the arrival of the so-called "New Growth Theory" (NGT) in the late 1980s and early 1990s.

From a theoretical perspective, these two renditions of the view that patents and legal monopoly are essential for innovation are what we have in mind when we speak of the currently received wisdom. Purely because of its very simplified analytical structure, we will use here the Aghion and Howitt (1992) model as our representative stand-in for this otherwise vast literature. Almost everything we say applies equally well to all other NGT models produced since the middle 1980s (in particular to those of Gene Grossman and Elhanan Helpman, and of Paul Romer, which were both prior to that of Aghion and Howitt) as well as to the classical contributions in the field of industrial organization.

This literature posits a tradeoff between "static efficiency," achievable under free competition insofar as this leads to an efficient allocation of resources for a given technology, and "dynamic efficiency," which is due to technological progress driven by patented innovations aimed at acquiring a monopoly. While competition may yield static efficiency, it fails to deliver the dynamic one; because the latter is, arguably, more important than the former, free competition is not good for society. Patents are therefore a "necessary evil": even if they sustain a technologically artificial monopoly power, thereby reducing consumer surplus while they last, they are a necessary tool for fostering dynamic efficiency over time. Innovative efforts would find no reward if the monopoly power that patents create were absent—no patents, no party.

There are two key technological assumptions underlying the prediction that, under competition and free entry, dynamic inefficiency (i.e., suboptimal innovative efforts) would be the outcome:

- The private fixed costs of innovation are ''large,'' at least relative to the size of the market and the evaluation (willingness to pay) that the marginal consumer would give of the new good.
- Imitation is both simple and cheap, so much so that, within a small span of time, so much productive capacity can be built up that every competitor would be led to play a Bertrand game when it comes to pricing.

It is obvious that, in a world where these two assumptions are simultaneously realized, one would observe no innovation absent patent's protection. Innovators would know that, as soon as they had sunk their fixed cost and introduced the new good or the new production method, imitators would flock into the market at zero cost, building up excessive capacity and, because of Bertrand competition, driving prices to variable marginal costs. Because the latter makes it impossible to cover the initial fixed cost, entrepreneurs would not even bother trying to innovate, thereby leading the economy to technological stagnation. From this reasoning there follows, on the one hand, the prediction that industries where imitation is easy will not see much technological progress and, on the other, the prediction that, in the presence of strong patents, innovation will instead thrive. This is statement (a) in the introduction.

It is worth stressing that statement (b) is neither marginal nor dispensable in this framework. When (b) fails, the model predicts that (a) also will fail and innovation will come to a halt even in a world where patents guarantee innovators a strong monopoly power. To see this, assume the current incumbent monopolist can—either through a sequence of very small innovations or by purchasing the key factors its competitors need in carrying out their research and development efforts—maintain its lead on the pack, thereby progressively reducing their chances to ever be able to ''jump over'' the incumbent with a breakthrough innovation. In other words, assume the incumbent monopolist can make valuable innovations extremely costly for its potential competitors; in this case, prediction (b) fails together with (a). In the Schumpeterian model, when the monopolist is so far ahead of its competitors to feel safe enough, the innovation machine stops operating and the economy becomes stagnant (see Piazza 2010 for an insightful formalization of this argument in a dynamic general equilibrium context). The monopoly-based theory of creative destruction does require monopolists to be ''destroyed'' every so often: if the monopolist never gets overtaken, then according to the theory, there cannot be any innovation.

3. CONCERNS ABOUT THE RECEIVED WISDOM: FACTS

Why should one doubt the sharp and reassuring predictions of such an elegant theoretical apparatus? In the Schumpeterian-NGT world, we can have our cake and eat it too: patented entrepreneurs are safely protected by the legal monopoly bestowed upon them, making higher and safer profits, while innovative activity (hence, technological progress and labor productivity growth) proceeds at the highest possible pace. This distinct characteristic of the Schumpeterian-NGT paradigm may well be the main reason behind its wide academic and public success. By telling us that monopoly power is socially benign because it generates ''dynamic efficiency,'' hence economic growth, at the very minor cost of some short-run losses of consumer surplus, it reconciles the desire of protecting what is already established with the urge to foster progress.

A world of low risk and high returns, though, seems too good to be true. At a minimum, it contradicts the basic economic intuition according to which there are no free lunches and, in a world of tradeoffs, rewards always entail some costs. More to the point, economic theory has been arguing for centuries that competition is conducive to good economic performances and good management practices, and all the available evidence (Nickell 1996; Blundell, Griffith, and Van Reenen 1999; Carlin, Schaffer, and Seabright 2004; and Okada 2005 are some of the main references within a very wide literature; for a decently recent survey, see Van Reenen 2010) suggests that, indeed, this is the case. If, in general, more competition has been shown to foster better allocation of resources, lower production costs, better management practices, higher workers' and entrepreneurs' effort, adoption of technological best practices, and so on, then it would be hard to believe on purely theoretical grounds that technological change and innovation should make such an extraordinary exception to this general and repeatedly confirmed rule.

This simple observation, after all, was at the root of an unfortunately little-known but remarkably prescient paper by George Stigler (1956), who—somewhat lonely—tried to argue that, to the best of his knowledge, most of the great innovations of any century had been created under conditions of competition and in the absence of patents and legally enforced monopoly power. That same paper contains a simple, somewhat artisanal empirical test in which Stigler uses pre–World War II data to regress a raw estimate of sectorial labor productivity growth against an index of sectorial concentration, finding that, indeed, the less concentrated an industry was, the faster labor productivity grew, at least back then. As far as we can tell, that was the first empirical refutation ever published of what is now (and was becoming then) the established wisdom. Times may have changed since (and they have indeed changed), hence the need for replicating Stigler's test using more recent data and more advanced statistical techniques, which we do in Section 5.

Before getting to it, let us consider a few other facts casting serious doubts on the realism of the Schumpeterian-NGT theory.

3.1 Who Innovates?

A key prediction of the Schumpeterian view of innovation is that the latter is the special product of large business organizations—their unique contribution to economic growth and to the increased social welfare it brings about. As a matter of fact, Schumpeter's argument rests more on the deep pockets and planning capabilities of large conglomerates than on patents and intellectual property per se. After convincing himself (do not ask us how) that socialist planning was the best way to achieve economic growth, he was intent on convincing his readers that through the actions of the protected monopolies, we could have our cake and eat it too: socialist planning for workers and consumers, and private property for the few lucky monopolists. Unfortunately, the historical and statistical evidence points to the exact opposite of what Schumpeter predicted. While large corporations are often (not always, maybe not even most of the time, but certainly often) the outcome of major innovations—sometimes patented and sometimes not; think of Ford and Edison, Sears and Amazon, Microsoft and Google, AT&T and Wal-Mart—major innovations are seldom if ever the product of large and monopolistic corporations. On the contrary, breakthrough innovations are more often than not the product of small firms—competitive and creative outsiders that must compete with established incumbents and are able to do so because of their creative superiority. This fact, which has been remembered and then forgotten innumerable times in the applied economics literature, is documented once again in the recent work of William Baumol, one of the most articulate supporters of the central role played by the ''creative destruction'' mechanism in generating economic growth. So, for example, in Baumol (2010, chap. 2) we find an excellent summary of a series of studies, produced between 1995 and 2004 by the U.S. Small Business Administration, documenting that, in spite of the fact that the bulk of R&D activity is carried out by large corporations, the lion's share of technological breakthroughs come from small firms. One may argue that these findings are biased by the fact that the SBA's task is to argue exactly that. This is a correct observation that, nevertheless, cannot erase the fact that the data reported in those studies have withstood various years of attempted criticism and that the author quoting them with approval is one of the strongest supporters of the Schumpeterian view of technological progress. A second, critical observation may point out that a number of large corporations (IBM, Xerox, Bell Labs, or Microsoft) have been innovating a lot. Leave aside the fact that one would have a very hard time making the list longer than the one above; the fact is that the first three—IBM, Xerox, and Bell/AT&T—came up with lots of good ideas when they were the dominant monopolists in their industries but left those ideas in their drawers, not turning them into actual innovations until the breakup of their monopoly power forced them to compete with new entrants. Only then did their labs' good ideas turn into actual innovations, which is exactly the opposite of what the Schumpeterian's viewpoint implies. As for Microsoft, the answer is even simpler, if historically symmetrical to the previous one. Microsoft *was* a great innovative company until it had to fight to stay ahead of its competitors, but it stopped innovating once it acquired its monopoly power. As a matter of fact, Microsoft would already be history if it were not for the monopolistic grip it maintains on the market through its Windows operating system.

Establishing that the bulk of innovations does not come from large corporations but from small ones may or may not matter for those subscribing to a more ''elastic'' view of the creative destruction mechanism. But it should at least serve as a reminder that, if patents are socially helpful, it is not because they allow the big conglomerates to innovate (as in Schumpeter's initial theory) but—maybe—because they allow the newcomers to enter a market. Hence, as far as evidence is concerned, we are restricted to considering the hypothesis that patents are good mostly for small firms and are instrumental to the inception of new industries. Is this the case?

This leads to a more sharply defined empirical question. Consider the 10 or 20 most economically relevant ''new industries'' of the last century or two: from electrical machines and appliances to chemical and then pharmaceutical products, from the car to the aviation industries, from software to hardware to mobile devices, from retail distribution to steel, from the movie and television industry to banking and finance. Most of these sectors are now relatively mature and dominated by a few large companies, either country-by-country or even worldwide, though to different extents. Nevertheless, when we look back at their inception, be it more than a century or half-century ago, not only do we not see any big conglomerate ''inventing'' (say) the personal computer or the motor-car and protecting it with a patent, but—practically always—we see groups of small and unprotected entrepreneurs introducing the new goods or services without much patent protection, at least in the very initial stages of the industry. The fact is that—even in the few cases when the innovation that revolutionized an industry comes from within pre-existing and relatively large companies in the same industry (think of the role that Wal-Mart has played, for example, in dramatically increasing the productivity of the retail and distribution sector, or that the now-all-defunct investment banks played in the 1970s and 1980s in creating the modern financial industry)—patents played either a minor role or no role at all in the gestation of the initial innovation!

In the light of these and plenty of similar observations, we wonder why in the applied industrial organization literature the following question seems to have never attracted the attention of researchers: what does actually happen over the life cycle of a new industry? The question is, at the same time, historical and statistical—purely descriptive in some sense, but of great theoretical relevance. How did the new industries that have actually emerged since the inception of the industrial revolution about two centuries ago come about?

How many were fostered by a preexisting large conglomerate and/or protected by an array of patents? Along the life cycle of a new good/technology, when does intellectual property enter the scene and start playing a (positive) role in fostering innovation? Is intellectual property there from the very beginning? Is it really the essential obstetrician without whom the healthy infant would never see the light of day, let alone grow to become the powerful athlete it eventually will? Or is it something that comes much later in the life of innovative industries, when the rate of innovation slows down, the technology tends to mature, and the classic ''shake out'' stage is under way or has just passed? Our guess, maybe anecdotal but not completely uneducated, is that the ''stylized fact'' emerging from such a careful historical investigation would be exactly the one we are suggesting here through our rhetorical questions. As far as we can tell, the main role intellectual property has played in the initial development of the great industries of the last two centuries has been either none or to delay some of them—see, for example, Boldrin and Levine (2008a, chaps. 1 and 8) for the cases of the steam engine and the airplane, respectively. In the few cases in which patents helped some among the initial innovators, such help was either damaging to the growth of the industry overall (as in the case of radio through the controversial Marconi patents), or it played an ambiguous role by first enabling either ''theft'' or something pretty close to it and then leading to the creation of long-lasting and scarcely innovative monopolies (as in the cases of Alexander Graham Bell and the telephone, and of RCA and the television). The only important exceptions to these stylized facts we can think of are associated with the name of Thomas Edison—the light bulb and related electrical equipment.

Most certainly we may be wrong, which is why we hope someone better equipped than we are will find the time and energy to look carefully into this issue. For the time being, though, all the modern industrial history available in print and online says that the Schumpeterian-NGT prediction is much more a fable than history.

3.2 Revolving-door Monopolists?

Next is the observation that, while in the established model of creative destruction progress comes from a new (patent-protected) monopolist overcoming the incumbent one with a superior innovation rendering the latter's old patent irrelevant, this seldom happens in practice.

This observation readily applies to mature industries, where we observe two kinds of long-run market structures emerging. Most common is the case in which the efficient technology is so widely accessible that the sector has remained competitive and still experiences a large amount of entry and exit even many decades after having reached its maturity—for example, textile and apparel, furniture and house appliances, food and drinks, retail distribution, restoration and hospitality, and so on and so forth for a long array of ''traditional'' goods and services. These mature industries still experience a fairly large amount of creative destruction, with local or temporary leaders emerging and disappearing at a relatively high frequency and, more importantly, a sustained growth of labor productivity. Still, all this is achieved with little or no use of patents as a tool for enforcing monopoly power, thereby contradicting at least one of the two basic predictions of the Schumpeterian-NGT model (innovations keep taking place in these industries without the help of patents). The same model is also contradicted for the complementary reason by those mature industries that evolved toward a long-run oligopolistic market structure and in which patents are frequently used to enforce monopoly rights over innovations. In these cases—for example, cars and trucks, trains and airplanes, personal computers, proprietary software applications, shipbuilding, medicines, durable goods in general—market leaders tend to remain the very same few for very long periods of time (decades and decades, in fact) with very little, often literally zero, entry of creative outsiders. In all these industries, patents are certainly used and a more-or-less high level of innovative activity does take place, but the creative destruction mechanism predicted by the Schumpeterian-NGT model never materializes. There may be ''creation,'' but there is very little destruction: after a couple of decades, Microsoft Word—no matter how cumbersome and little different from 20 years ago it happens to be—still is the market leader in word processing, not to speak of Excel. The last time the market for word processing (or for spreadsheets) experienced any kind of ''creative destruction,'' patents were *verboten* in the software industry. This is not an exceptional case, as a quick examination of the industries listed above (or of your own consumption basket, for that matter) will easily confirm; there is a lot of healthy creative destruction in free-market economies, but little of it comes from the new creative patentee overtaking the old, and no longer creative, patentee.

3.3 More Patents = Higher Productivity?

Let us move on to a more delicate issue: Is there a sharp and well-defined one-to-one relationship between patents and actual technological innovations? No matter how you turn it around, the Schumpeterian-NGT theory predicts that the stronger the monopoly power and the higher the (impact-weighted) number of patents issued, the stronger productivity growth should be—without "if" and without "but." Coherent with such a prediction, patents, their numbers, and the number of times they are cited by subsequent patents are by far the most commonly used measures of innovation in most empirical industrial organization studies. Apart from the predictions of the received wisdom, this common academic habit has another reasonable justification: plenty of high-quality econometric work (starting with Pakes 1986) has shown that citation indices are a good measure of patents' quality and market impact. They are, for example, important predictors of the stock market valuation of the firm owning the patents. Frequently cited patents are economically very valuable, while scarcely cited patents tend to be associated with valueless innovations. In going from this finding to the conclusion that patent citations are an overall good measure of innovative activity, though, there is a far-from-obvious logical step, which we consider next.

The problem is twofold: It seems reasonable to believe that if a given innovation is both valuable and patentable, the innovator will have an incentive to seek a patent. While a number of surveys (Levin et al. 1987; Cohen, Nelson, and Walsh 2000) show that innovative firms very often rely on tools other than patents (e.g., secrecy, lead time, and so on) to achieve and maintain their competitive advantage, there is no doubt that, when possible, firms do patent their valuable innovations. It follows, therefore, that some innovations are contained in patents and that some patents contain actual innovations. The problem arises once we try replacing the two occurrences of the word "some" in the last statement with the words "almost all." Safely performing such substitutions is necessary for treating "patents and patent citations" as adequate empirical measures of "technological change."

In very many industries, patenting was not even an option until a short while ago (e.g., the 1970s for plants, the 1990s for software, the 2000s for business practices, etc.) and it is still not a truly usable option for many other ones (e.g., retail and distribution, financial products, fashion and design, telecommunications, and so on). What this means is that lots of valuable innovations are not captured either by patents or by patent citations, making the latter an incomplete measure of the former. As the data below show, a very large number of productivity-enhancing innovations are most certainly not captured by patents. The other side of the problem is subtler: how many patents are issued (consequently, how often a patent is cited by subsequent ones) is determined by the specific legal and industrial organization characteristics of each different industry. Tens of thousands of patents are awarded each year in the software industry, while those in the car industry are counted in the hundreds. Some of these patents have innovative contents, but many others are purely ''defensive'' patents taken out for purely legal purposes and with little or no innovation inside them. The cumulative research work of James Bessen and his coauthors (e.g., Bessen and Hunt 2003) has made this fact plain in the case of the software industry.

As patenting became legally possible and spread widely across the industry, the actual amount of innovation carried out certainly did not increase and probably diminished, thereby making patents (and citations) a very poor measure of what we often take them to be. Our own somewhat artisanal examination of the agricultural sector (e.g., Boldrin and Levine 2008a, chap. 3) has led us to the same conclusion. Since they were legally allowed in the early 1970s, the number of agricultural patents has exploded—especially in recent decades thanks to the introduction of bio-engineered products—but there is no evidence of a structural break whatsoever in the time series of agricultural TFP either in the United States or in Europe. Hence, whatever it is that the dramatic increase in agricultural patents may be measuring, it cannot be the movement in agricultural productivity. Because it is the latter we care about, one should conclude that neither almost all innovations are contained in patents nor almost all patents have innovative content.

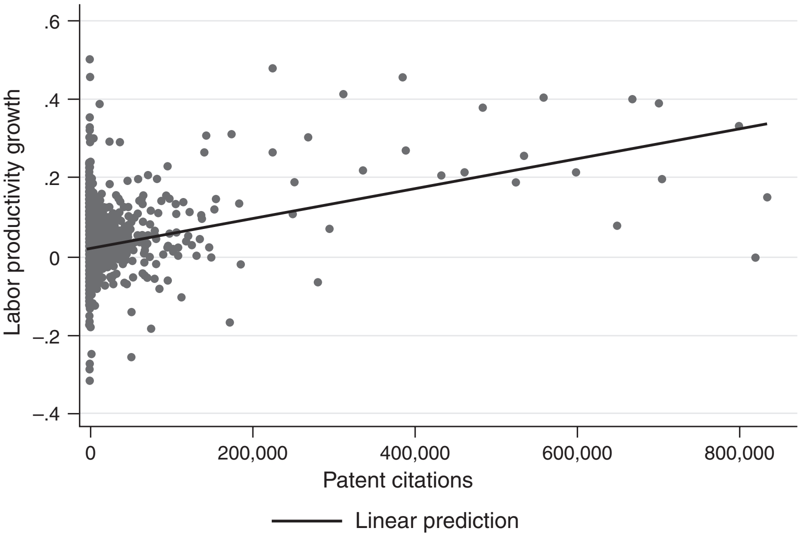
Case studies aside, how serious is this problem, statistically speaking, across a wide array of sectors and firms? In other words, for the overall economy, how poor a measure of productivity growth are patents and their citations? We try shedding some light on this question by computing conditional and unconditional correlations between patents and productivity measures from the same datasets we use later in Section 5 to carry out our main tests. In this instance, we have computed measures of labor productivity and TFP, sector by sector, together with patent numbers and citations for the firms in each sector. Next, we have regressed both their levels and growth rates on each other in a variety of reasonable combinations, both unconditionally and conditioning for investment activity in previous years and a number of common-sense control variables.

While not perfectly homogenous, our findings are nearly so. Their basic message is that, except for a couple of specifications reported below, there is not even a significant positive correlation between patents and productivity. This is a more surprising result than one would predict on the basis of purely theoretical considerations. In fact, one would expect patents to be at least a decent predictor of productivity growth across sectors—certainly for the last couple of decades, during which their use was extended to more and more sectors. Instead we find a weak correlation in only one of the many possible specifications, and no correlation at all otherwise! The data suggest, in other words, that the use of patents either as a defensive or as a rent-seeking tool (Boldrin and Levine 2004, 2008b) is actually more widespread than we had estimated it to be.

This specification yields a statistically significant, but very small, positive coefficient when we apply it to U.S. Bureau of Labor Statistics (BLS) output data, and a coefficient not statistically different from zero when National Bureau of Economic Research (NBER) output data are used. A quick look at Figure 1 suggests that a few data points may be driving the result, hence Figure 2 reports the same data without the top 5 percent of observations for citations. In Figure 2, the correlation is gone.

The role that the three statistical outlier sectors play in satisfying the received wisdom's predictions is revealed by computing the 20-year mean, at the North American Industry Classification System 4-digit level, for both variables in each sector. This is depicted in Figure 3. Also, for sectorial averages, once the top 5 percent values of the ''citations'' variable are dropped, the statistical correlation weakens substantially and, also in this subsample, appears to be determined by a small number of high-citation sectors. This is depicted in Figure 4.

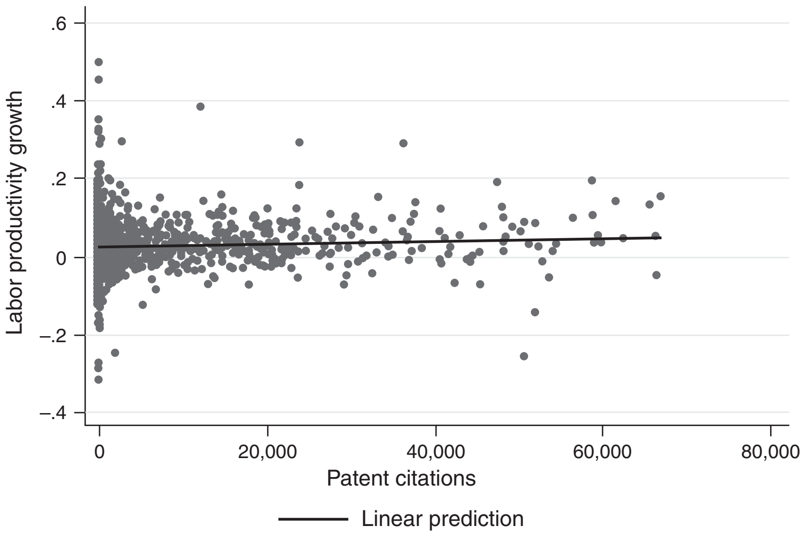
Figure 1
Patent Citations vs. Labor Productivity Growth
(1987–2008)



Very similar (in fact, weaker) results obtain when measures of TFP are used. In the Appendix, we report a number of additional cloud-like plots that are representative of the kind of results one obtains with different specifications. As the various graphs vividly illustrate, even in the very long run, patents and their citations measure only vaguely those innovations that actually increase either labor productivity or TFP, and do so only for a handful of sectors. Most observations are concentrated near the origin of the horizontal axis, corresponding to very few or no patents and patent citations at all; they display as much variance in the measured growth rates of productivity as the overall sample does, signaling that the larger fraction of productivity-enhancing innovations are not captured by patents.

This suggests that patents should not be used to assess which firm innovates more and which firm innovates less, unless one is capable of controlling for industry-specific effects and the various strategic and legal considerations different firms face at different points in time. From the point of view of economic theory and social

Figure 2
Patent Citations vs. Labor Productivity Growth
(Top 5 percent dropped)

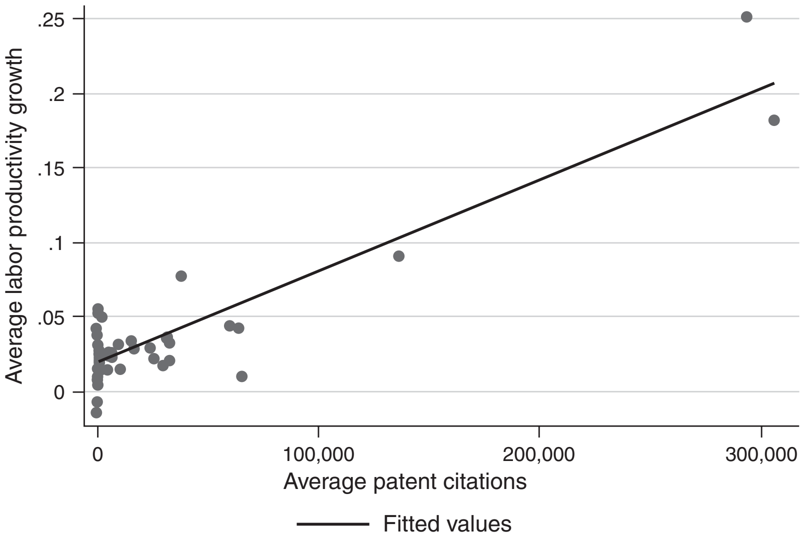


welfare, what matters is the growth of factors' productivity and not of the number of patents per se. Contrary to what it has been doing for decades under the influence of received wisdom, empirical research in this area should focus on the former as the proper measure of socially desirable outcomes, treating the latter as a tool—sometimes just a legal tool—that may or may not foster desirable innovations and productivity growth. Whether patents do or do not improve productivity should be the question we ask the data, not the preset answer from which we start our investigation.

3.4 Inverted-U Relation?

In recent years, an empirical ''middle of the road'' position has emerged within the Schumpeterian-NGT framework of analysis. This position is widely seen as a way of reconciling the theoretical predictions of that line of thought with the growing evidence that increasing monopoly power generates less, not more, innovation. Among others, the paper by Aghion et al. (2005) contains the main empirical findings giving support for this position, while the book

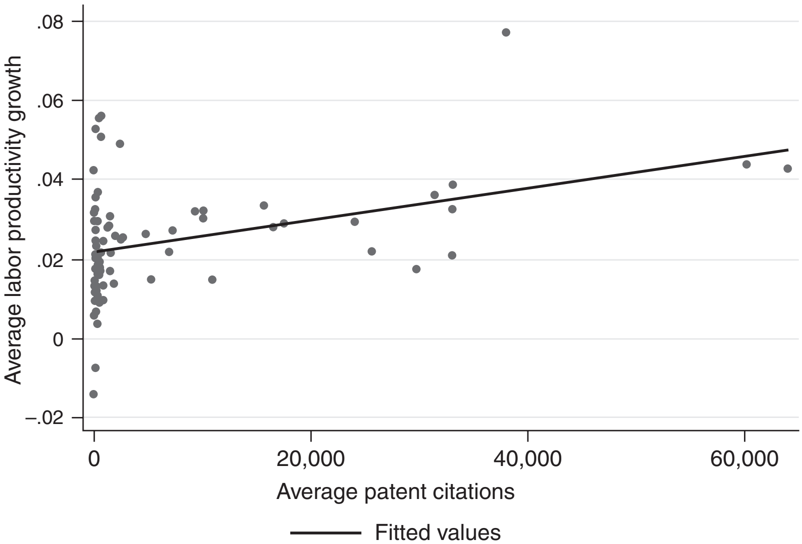
Figure 3
Average Patent Citations and Average Labor Productivity Growth



by Aghion and Griffith (2005) provides a broader overview of the whole line of analysis.

Aghion et al. (2005) develop a ''variety'' model in which competition is measured by the elasticity of substitution within pairs of goods, each pair produced by a duopolistic industry. In this setting, the higher the elasticity of substitution between the two goods is, the higher the return from innovation for either of the two duopolists will be. Assuming that sometimes one of the duopolists finds itself to be a technological leader relative to the other, a high elasticity of substitution between the two goods reduces the incentives to innovate for the laggard when its distance from the leader becomes particularly large. The authors interpret this kind of model as predicting that the maximum innovative effort will obtain at some ''intermediate'' position between perfect substitutability (competition) and perfect complementarity (monopoly). They compare these predictions with the patenting activity in the United States for a panel of United Kingdom manufacturing firms, claiming that their

Figure 4
Average Patent Citations and Average Labor Productivity Growth
(Top 5 percent dropped)

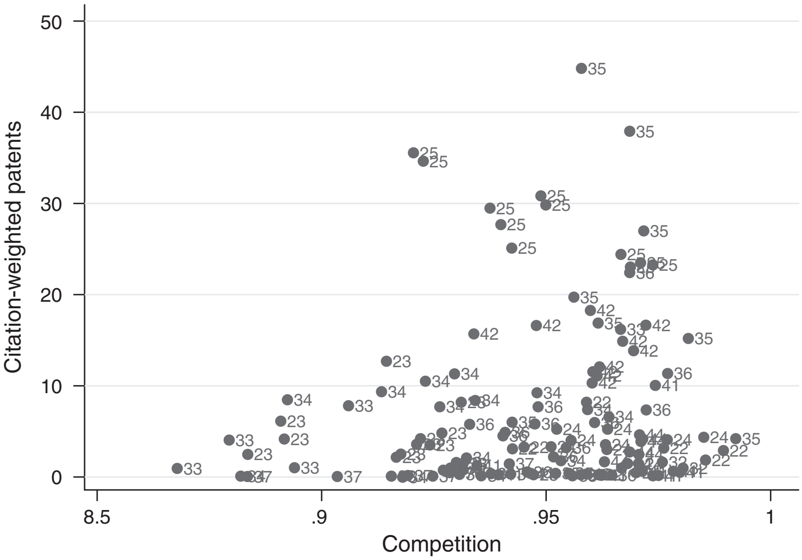


model's predictions are supported by the data. This is interpreted as a vindication of the basic intuition behind the Schumpeterian-NGT theories, according to which at least some degree of patent protection and monopoly power is good for technological progress. *In medio stat virtus.*

The robustness and even the very same existence of such empirical findings are questioned in the works of Correa (2010), first, and then of Hashmi (2011). The latter re-examines the inverted-U relationship between competition and innovation by using a different dataset covering publicly traded manufacturing firms in the United States between 1976 and 2001. Apart from a minor statistical detail, Hashmi mimics the estimation strategy of Aghion et al. (2005). He replicates their results for the UK dataset but, when using the U.S. dataset, he finds a robust positive relationship between competition (as measured by the inverse of markups) and innovation (as measured by citation-weighted patents). Hashmi conjectures that the different findings may be due to the underlying characteristics of the two different economies. In particular, he claims that a possible explanation of the sharply different results may come from the fact that (in his view) the U.S. manufacturing industries are technologically more neck-and-neck than their counterparts in the UK. In other words, there is a lot more competition among firms in the United States than in the UK. In this case, because no competitor is ever ''far behind'' the leader, the incentives to innovate that competition induces are stronger; because there are not many laggards giving up the race, there are also very few leaders that are able to keep the lead without further innovations. The argument is more than a bit lopsided because ''competition,'' in these regressions, means a low markup, and the argument we just summarized does not explain why, in the United States, firms with very low markups should innovate a lot while in the UK the same kind of firms innovate very little. Further, as Hashmi (2011) remarks, this conjecture is not really supported by the data, and there is no independent evidence in the literature that this should be the case. Finally, we note that, even if the conjecture were supported by the data, it would simply mean that, indeed, the more competition there is, the better competition works. This tautology, at the very end, does not account for the different empirical results between the U.S. and UK firms, which remains a puzzle.

The investigation carried out by Correa (2010) shows, instead, that there is no puzzle and that this particular conjecture is not necessary to account for the different correlations observed in the two datasets. This is because, most likely, the inverted-U pattern that Aghion et al. (2005) claim to have discovered in the UK data is either not robust to very reasonable perturbations or it is not there at all. Guided by the well-documented recent history of the U.S. patent system, Correa begins his investigation by carrying out a simple Chow test on the Aghion et al. (2005) data to find evidence that a structural break took place in the early 1980s. This structural break coincides with the establishment of the United States Court of Appeals for the Federal Circuit (CAFC) in 1982. Scholars of intellectual property have amply documented that the establishment of the CAFC had a dramatic impact on the enforceability of patents and greatly strengthened the position of patent holders, see, for example, Jaffe and Lerner (2004), Kortum and Lerner (1998). It therefore increased the incentive to apply for patents, and those applications started to

Figure 5
Competition and Patent Citations Prior to the CAFC
(1973–1982)

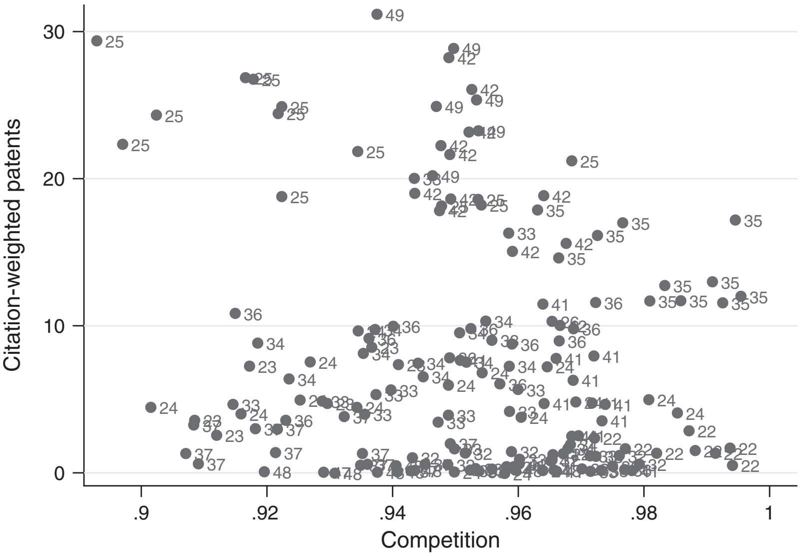


increase at about that time. By taking into account this break, Correa finds a positive relationship between innovation and competition during the period 1973–1982 and no statistically significant innovation-competition relationship during the period 1983–1994.

Figure 5 displays the sample observations of innovation and competition levels before the establishment of the CAFC (period 1973–1982). We can see that some particular industries, such as Motor Vehicles (Standard Industrial Classification 2-digit code 35) and Chemicals (SIC-2 25), have higher levels of innovation than the rest of the manufacturing sector. However, as Figure 6 shows, after the establishment of the CAFC, both the Motor Vehicles and the Chemicals industries decreased their innovation intensity, while industries such as Other Manufacturing (SIC-2 49), Sugar Beverage and Tobacco (SIC-2 42), and Office and Computing Machinery (SIC-2 33) increased their number of citation-weighted patents.

The Chow test looks for only one structural break without using additional information about its location in time. Clearly, the break

Figure 6
Competition and Patent Citation after the Creation of the CAFC
(1983–1994)



could have taken place in another year or there could be more than one break, hence a sup-Wald test is performed, which makes no assumptions about either the number or the location of the structural breaks. This test finds just one structural break and places it in 1981, while the CAFC was established only in 1982. However, it is worth noting that in the dataset, a patent is assigned to the year in which the application was filed rather than the one in which it was granted. While it is now greatly increased, reaching almost three years, a substantial lag between applications and grants existed already in the early 1980s. Furthermore, the establishment of the CAFC did not come as a surprise to anyone but was the consequence of a long and public debate about the opportunity of introducing such a court to strengthen the U.S. patent system. It seems therefore reasonable to expect patent applications to react somewhat earlier on, in anticipation of the October 1982 establishment.

We leave it to the reader to decide which way the final jury should lean. We are satisfied with having shown that, indeed, once one takes into account a little tiny bit of well-known history, the inverted-U relationship is gone, and a clear upward-sloping relation connects the degree of competition with the strength of innovation also in the particular dataset used by Aghion et al. (2005). That is, we like to recall, the one and only dataset discovered so far in which a modified version of the Schumpeterian-NGT theory appears to be vindicated.

4. A DIFFERENT POINT OF VIEW

In the data, one can hardly find any support either for the claim that stronger intellectual property regimes favor technological change or for the general theory according to which monopoly power fosters innovation. (There is, in fact, good evidence for the opposite.) There is scant evidence even for the claim that patents are a reliable measure of productivity growth. Thus it is important to understand how markets for innovation and technology adoption might function in the absence of intellectual property. Innovative activity has taken place historically and does take place currently, as plenty of evidence shows, in markets where intellectual property restrictions are either absent or even not allowed. Hence, in order to explain the existence of something that standard theory predicts should not even exist (i.e., competitive innovation), one needs to develop an alternative theoretical framework, which we briefly outline next.

In the paper we quoted earlier, Stigler (1956, p. 274) argues that monopoly is completely unnecessary to provide incentives for innovation.

> There can be rewards—and great ones—to the successful competitive innovator. For example, [consider] the mail-order business.... The innovators ... were Aaron Montgomery Ward, who opened the first general merchandise establishment in 1872, and Richard Sears.... Sears soon lifted his company to a dominant position by his magnificent merchandising talents, and he obtained a modest fortune, and his partner Rosenwald an immodest one. At no time were there any conventional monopolistic practices, and at all times there were rivals within the industry and other industries making near-perfect substitutes.

In more recent times, Liebowitz (1985), Hellwig and Irmen (2001), Quah (2002), Legros (2005), and Boldrin and Levine (1997, 2005, 2006, 2008a,b) have all examined the competitive rents that accrue to innovators due to ''limited capacity''—the fact that, in a competitive market, the owners of a fixed factor (first copy of an idea) are the recipients of all downstream rents originating from it, and that an infinite number of copies cannot be made instantaneously. The conclusion is that innovation will take place even without intellectual property, as it often has in the past—see, for example, the cases mentioned by Moser (2005). While some of this work shows that there may be too little innovation under competition due to the indivisible nature of the initial copy of ideas, it also suggests that the appropriate remedy is unlikely to be a government-granted monopoly.

On the one hand, it seems transparent that providing a reward for innovation in the form of a monopoly can only increase innovation, at least on impact. Indeed, this idea is so obvious it seems to have an intellectual grip on the economics profession that a half-century of empirical evidence to the contrary has been unable to break. There is, however, a flip side to a patent system as opposed to an individual patent: a patent system not only rewards innovators, it also makes it more costly to innovate in the face of the many licenses that need to be acquired in order to bring a new product to market. Hence, from a purely theoretical point of view, the impact of patenting on innovation is ambiguous: it both encourages and discourages innovation. Once this fact is taken into due account, it is perhaps not such a surprise that no evidence has emerged that there is a net positive effect. Scotchmer (1991) is among the first to point out the problem that occurs when innovations build on existing ideas. Formal models, first in the form of a simple example and later in the form of a detailed model of innovation, were provided in Boldrin and Levine (2005).

On the positive side, Boldrin and Levine (2006) examine innovation that is driven by the ''need'' for new technologies when old technologies have exhausted their growth potential. In such a setting, perfect competition delivers the first-best, as competitive rents are exactly what is needed to provide the socially optimal level of invention. Hence, any intervention in the form of a government-granted monopoly strictly reduces welfare because a monopolist would accumulate each technology more slowly than under competition, thereby increasing the time it takes to move from one technology to the next, more productive one, thereby reducing the long-run growth rate of consumption.

A further aspect of the theoretical problem, as Stigler briefly noted in his 1956 article, is that the formal definition of competition is capable only of capturing the final, stylized outcome and fails to describe the process. In a (perfectly) competitive industry, the participating firms are small relative to the size of the market, produce the same identical goods with the same identical technology, and act as price-takers because there is nothing they can do that can alter anything of relevance in the market in which they participate. Such a description of "competitive outcomes" may be useful for the study of general equilibrium arrangements in certain settings, but it hardly seems capable of describing the underlying intuition according to which "competing entrepreneurs" try to outsmart each other by both imitating the best practices and improving upon them, reducing production costs as much as possible (thereby de facto implementing technological change) to finally bring about some form of cost/price equalization among the surviving firms. It is the latter, though, and not the former that one has in mind when saying, intuitively, that "competition fosters innovation."

Is "monopolistic or imperfect" competition, then, the proper analytical answer to this unpleasant state of affairs? While one could be tempted to say "yes," at least instinctively, and then follow along the lines of Aghion et al. (2005) and identify the extent of competition with the degree to which one good is a substitute for another, there are good reasons not to go that way. The extent to which a new good is, or is not, a substitute for an already existing one is an endogenous equilibrium choice, not an institutional or technological parameter, as innovators elect which niche to enter on the basis of market incentives. This is, in fact, one of the channels through which competition fosters innovation: the free entry of imitators/innovators choosing either to expand total productive capacity or to build up on previous innovations by introducing close substitutes of the existing goods. Until we are able to build workable models of free entry in which this dynamic aspect of competition is captured over time, we must face the unpleasant choice of either going to the extreme of "perfect competition," in which everyone always does the same thing, or to that of "monopolistic competition," in which every firm is actually a monopolist acting in parallel to other, similar monopolists.

In any case, the key conceptual problem with the ''monopolistic competition'' approach to modeling competition is that it is grounded in the idea that imitation is costless, hence competitive rents are irrelevant and competition would always bring about Bertrand pricing. When imitation is costly but competitive rents are not immediately equal to zero—therefore, imitation is profitable—monopolistic competition evaporates, and the dynamic process of competition with free entry sets in. In practice, imitation *is* costly and, as noted above, a patent system serves only to discourage downstream imitation. The latter point is important because in practice—as shown by the managerial surveys cited earlier—innovations tend to be protected more by the difficulty of imitation than through government intervention in the form of patents. This means that, de facto, the role of patents is to discourage competition, free entry, and indeed innovation; old incumbent firms acquire large patent portfolios covering tens of thousands of known ideas with the aim of using them as barriers against any entrant that does not have a similar portfolio.

Finally, let us consider the following often-heard criticism of the theoretical argument according to which competition favors innovation more than (legal) monopoly does:

> The Boldrin & Levine model has a built-in first mover advantage. Of course, then, if the first mover advantage is big enough you do not need patents. I never understood why we should pay attention to such a trivial point (theoretically, the empirical question of whether first mover advantages would be enough is of course relevant). Is there something that I have been missing all of these years?

It is true that the essence of the Boldrin and Levine model is a first-mover advantage generating competitive rents. Boldin and Levine repeatedly mention it in their original paper on perfectly competitive innovation (Boldrin and Levine 1997), which was in fact meant to show how the traditional Edgeworth-Marshall framework of a competitive industry with free entry and minimum plant size could easily be adapted to model repeated innovation in a growth-theoretical setting. But there are two aspects in Boldrin and Levine's paper that are not trivial at all, at least not for the modern innovation and growth literature. First, they make the argument that there is always a first-mover advantage and that one can predict how large it is on the basis of observables. In other words, they give some structure to what the above criticism calls an advantage ''big enough'' and show under which conditions it may or may not obtain in practice. For instance, given the elasticity of demand, they show that the first-mover advantage (and the incentive to undertake an innovation) depends on two things: the initial cost of getting the first unit of output (the prototype), and the rate at which the invention can be copied. Second (and related to the previous point), even in markets where it is very easy to reproduce the innovation, the rent of the first mover can potentially increase if consumptions between periods are substitutes.

Actually, we think there is a deeper point that the expression ''first-mover advantage'' misses. Also, in a regime of monopolistic competition, there is a first-mover advantage—a short-run monopoly. But whether the first-mover advantage is a legally induced monopoly power (i.e., a market distortion) or competitive rent (fully efficient) is quite relevant for both positive and normative analysis. It is relevant for positive empirical analysis because one can use statistical data to quantify the two uses of ''enough'' in the statement, ''[W]hen the initial fixed cost is small enough relative to the size of the market and the cost of copying is high enough, we will observe sustained innovation under conditions of competition.'' It is relevant for normative analysis because, to the extent it allows us to explain the thousands of episodes of competitive innovation observed in reality, it provides us with guidance for setting policies in this area. Second, the expression ''first-mover advantage'' seems to imply that, once there is a ''second mover,'' the first-mover advantage goes away. It does not, and the framework mentioned earlier captures this important fact well, while the one of monopolistic competition misses it completely. This is because in the Edgeworth-Marshall framework, one assumes that capacity is bounded at every point in time and costly to accumulate, which is not the case in the monopolistic competition case. Because of this assumption, in a competitive industry, rents are generated more or less for the entire life of the industry; in the baseline case, without any external effects, all the rents do go to the first mover, but in more general cases, they also accrue to imitators and downstream innovators. We do not think that is what has been historically meant by ''first-mover advantage,'' and we would be pretty surprised if the above criticism meant that. As far as we can tell from the strong resistance this set of propositions still faces more than a decade after they were first published, these statements are not widely accepted. The propositions may be wrong, but they are definitely not ''trivial.''

There is another point in which the reported criticism is perfectly correct. Ultimately, it is an empirical question whether absence or presence of patents distorts markets more. The ''this is trivial'' criticism seems unaware of a very long literature in and outside of economics asserting that this is not the case and that, as a matter of theory, absence of patents distorts markets more than their presence. As we have documented, there is an ever-growing, long list of empirical economists surprised to learn that, in the data, patents lowered rather than raised innovation, and they have no idea why that might be true. The theory of competitive innovation provides a testable answer to this long-standing puzzle.

4.1 The Embodiment Issue

The ''embodiment hypothesis'' is a crucial step in the theoretical argument claiming that competition, in the sense of free entry and the right to imitate, fosters innovation and economic progress. Before moving forward to consider recent microeconomic evidence supporting our claim, we should briefly explain why the embodiment/disembodiment controversy is relevant in the context of our research.

First off, what is the embodiment/disembodiment controversy? It centers on the fact that technological advances may or may not be obtained without embodying them in something material and expensive to either produce or acquire—that is, in some ''capital,'' be it physical or human or organizational. Traditional treatments of the TFP measure describe the latter as completely disembodied and unrelated to investment expenditure. Most likely, this did not correspond to the intuition many researchers had of the nature and causes of TFP—certainly, it was not what Robert Solow had in mind. Nevertheless, it was the formalism adopted in writing $Y = AF(K, L)$, where Y is output, A is TFP, K is the stock of capital, and L is the flow of labor entering the neoclassical production function F. The same is true for the literature we have classified as Schumpeterian-NGT:

while some of those arguments would, formally, go through even if the new technology were embodied in some kind of capital good, most models use a disembodied representation of technological progress to formalize their argument. There are purely technical reasons for this choice (i.e., deriving a balanced growth path), but there are also intuitive and conceptual ones, which should be understood, as they are relevant for the issue concerning us here.

The fundamental link between disembodiment and the Schumpeterian-NGT view of the world is related to the assumption of unbounded productive capacity that the Schumpeterian-NGT view needs to make in order to reach its specific conclusions. In fact, for the ''imitation will destroy innovator's rents'' argument to work, one must assume that imitators will force Bertrand pricing almost immediately. For Bertrand pricing to make sense, one needs a situation in which each of the competitors is capable of satisfying total demand all alone, which requires productive capacity to be instantaneously expandable. While one may pretend that the available quantities of physical and human capital of a certain kind (i.e., the kind embodying the new technology) can be increased at an infinite speed at no cost, this seems in clear violation of the basic economic intuition according to which there is no free lunch. In such context, the hypothesis that the new idea or technology is actually disembodied and that, therefore, one can make a very large number of copies of it at no cost and in no time seems necessary to move the argument forward. More generally, the view according to which ideas, once discovered, can be immediately copied by anyone at no cost requires, conceptually, a form of general disembodiment of the idea itself that, like a light in the sky, appears to, and is immediately usable by, everyone else, were it not for the legal restriction that patents imply.

At the opposite extreme, when a new idea is fully embodied either in an object owned by the creator or in her human capital, it is up to her to decide what to do with it, and unauthorized imitation becomes, if not impossible, certainly neither easy nor legal. This is the key fact allowing innovators to be rewarded for their inventions even in the absence of the specific privileges that intellectual property introduces. As a matter of fact, the theory of competitive innovation rests on the twin hypotheses that

- inventors have control of their inventions and will require appropriate payment to make them available to others, and

- imitation is always and everywhere costly because it requires either producing or acquiring the material object or the human capital embodying the innovation.

These two assumptions, which we consider most natural, imply that productive capacity embodying the new technology will build up only slowly, the price of the new good will be determined by demand for quite a while, and the position in which price equals average cost will be reached only in the long-run equilibrium of the competitive industry. As a consequence, competitive rents will be attained by both the innovator and her imitators. Also in this case, one can certainly conceive of circumstances in which such rents could obtain even when the innovation is completely disembodied, but those cases appear to be either exceptional or far-fetched.

In conclusion, the theory of competitive innovation rests on the fact that innovations are embodied while, symmetrically, the Schumpeterian-NGT view of innovation rests on the fact that innovations are disembodied. The latter is a factual issue, and it should be resolved empirically. Common-sense inspection of actual innovations fails to deliver (as far as we can tell) convincing examples of actual innovations that were completely disembodied and could be instantaneously reproduced in an unlimited amount with zero costs of imitation. Even the commonly used examples of ''digital goods'' (whose historical relevance is clearly limited, given that they were not around until a couple of decades ago) fail the test. Unless the original creator voluntarily releases the master copy of the digital good in question, making a large number of copies of it available to consumers is technically impossible, even assuming the machines through which reproduction and distribution of digital goods take place are costless, which they are not. In a world of free competition, creators will charge a price to release the master copy to imitators and, in equilibrium with free entry, it is easy to show that such a price will be equal to the net present value of all future rents (Boldrin and Levine 2008b).

The applied statistical literature on this argument is very large, and our reading is that—since Zvi Griliches' path-breaking 1957 Ph.D. dissertation ''Hybrid Corn: An Exploration in the Economics of Technological Change''—it leans almost completely in favor of the embodiment hypothesis. It should be pointed out here that Robert Solow himself, with whom many people tend to associate the idea of ''disembodied TFP,'' always squarely supported the view that the quantitatively relevant type of technological progress is, in fact, embodied in one form of capital or another.

In any case, this is not the appropriate place to engage in a full survey of the embodiment controversy, even if it may be the place to point out that an updated and careful summing-up of the half-century-old debate would be most welcome in light of its fundamental relevance when it comes to modeling innovation. We will therefore limit ourselves to summarizing the findings of a recent micro-investigation of this issue, which uses an original and very powerful dataset and that is one of the few to address squarely the questions that concern us, that is, the empirical relevance of ''disembodied external effects'' from technological innovation across competing firms.

Castiglionesi and Ornaghi (2011) make use of an unbalanced micro-panel dataset of Spanish manufacturing firms observed with annual frequency during the period 1990–2006. This dataset proves to be particularly suitable for disentangling the impact of specific individual sources of productivity growth, as it includes detailed observations on firms' outputs, inputs, proportion of skilled employees, types of capital investment undertaken, and modifications of the production processes. Moreover, a unique feature of this dataset is that it provides growth rates of firm-specific prices for outputs and intermediary inputs, thus allowing for the construction of a more reliable measure of firms' productivity change.

Their estimation builds up progressively from a simple regression that reveals a large and unexplained residual representing the (unweighted) average TFP growth across firms. Traditionally, this is taken as a measure of ''disembodied technological progress''; their goal is to demonstrate that, using the micro-observations listed above, one can show it is actually accounted for by very embodied investments in some kind of capital. To this end, they start by analyzing the contributions of traditional disembodied variables as sources of average TFP growth. They consider firm-specific learning-by-doing (LBD) and unpriced externalities such as those acting through aggregate human capital and the spillovers of R&D from firm to firm. The quantitative role played by such variables in accounting for TFP growth at the firm level is often argued to be strong evidence, possibly the strongest, in favor of the Schumpeterian-NGT theory of innovation and growth. Because, this theory says, technological progress is mostly disembodied and generates large externalities at no explicit cost to the recipients, patents are necessary to allow inventors to make their products excludable, therefore appropriating at least part of their social returns. Without patents, private appropriation of the returns from innovative investments would become impossible, and the value of the innovation for the innovator would dissipate through the external effects induced by free imitation. Castiglionesi and Ornaghi replicate in their dataset the well known earlier results showing that certain aggregate variables—interpreted as the source of disembodied technological progress and unpriced externalities—are correlated to average TFP growth at the industry level.

Next, they take into account the relevance of (human and physical) capital-embodied technological progress as an engine of TFP growth. They measure the impact of new capital goods on TFP by means of two variables: the average vintage of the physical capital and an index of new technology usage. They account for differences in human capital using two variables: firm wages and the percentage of R&D employees at the firm level. To deal with classical endogeneity issues, they estimate a specification with the ratio of skilled workers at the firm level instead of firm wages. Once the measures of embodied technological progress are considered, the variables that capture firm-specific LBD, human capital externalities, and R&D spillovers no longer show any relevance in affecting average TFP growth either at the firm or industry level. In other words, once they account for variations in physical and human capital that are measurable at the firm level, the external effects completely evaporate and are no longer relevant in ''explaining,'' even statistically, the movements in TFP. In fact, Castiglionesi and Ornaghi find that embodied variables alone can fully explain average TFP growth across firms and industries. Last but not least, in all specifications, constant returns to scale cannot be rejected.

Finally, in order to better assess firm-specific LBD and to look more carefully into the presence of potentially external effects, they consider two alternative measures that, arguably, are closer in spirit to the theoretical idea behind LBD: cumulative output since the introduction of a process-innovation, and time after the introduction of a process-innovation. These two variables should capture the idea that a change in the methods and techniques used for production must trigger a new learning cycle of the workforce at the firm level. When considered together with the embodied variables, these alternative measures of firm-specific LBD retain explanatory power. This is coherent with the classical definition of LBD: internal to the firm, short-lived, and due to the adoption of new processes. In other words, there is no evidence of unpriced spillovers in this dataset. These results, taken together with those relative to the embodiment of TFP, cast a long shadow on the idea that spillover effects external to the firm play a major role in technological progress and in the increase of TFP along the lines assumed in the Schumpeterian-NGT paradigm.

Summing up: Castiglionesi and Ornaghi's findings prove that things such as ''free imitation'' and ''disembodied technological progress''—generated via a chain of external spillovers from one firm to another—find scant support in actual microeconomic data. Innovations and technological change appear to originate at the firm level and are mostly embodied in investment decisions, which are therefore both costly and internal. Average TFP growth is fully explained by the kind of technical progress that is embodied in actual physical and human capital; economy-wide neutral (or disembodied) technical change plays virtually no role.

5. NEW FINDINGS

We have already seen in Section 3 that, contrary to what received wisdom keeps repeating, the available empirical evidence about the statistical relation between measures of competition and patents point to a positive link: more competition, more (highly cited) patents. Still, because we have also seen that patents and their citations are poor predictors of productivity growth, these results imply relatively little with respect to the policy issue at stake: Does competition foster productivity growth more or less than monopoly?

With the aim of replicating Stigler's test on more recent micro data, we have constructed a dataset that includes, among other measures, patent counts and patent citations for 220 4-digit SIC code industries over the period 1990–2001 and productivity growth for 85 4-digit NAICS code industries over the period 1987–2008. The raw data used to construct this dataset come from different sources. Firm balance-sheet and financial data available in Compustat are matched with firm-level data on patents retrieved from the NBER Patent Data described by Hall, Jaffe, and Trajtenberg (2001). Total output data for SIC-4 manufacturing industries are taken from the NBER and U.S. Census Bureau's Center for Economic Studies Manufacturing Industry Database described by Bartelsman and Gray (1996). Information on output, inputs, and productivity for NAICS-4 manufacturing industries are obtained from the BLS. Finally, we retrieve information on U.S. imports at SIC-4 and NAICS-4 industry level from the U.S. Department of Commerce and the U.S. International Trade Commission. The accounting data retrieved from Compustat include sales S (item 12), gross capital K (item 7), operating profits OP (item 13), and advertising expenditure A (item 45) for the period 1990–2006. The data we use refer to all firms in the manufacturing sector; this includes 7,432 firms divided among 220 industries according to the 4-digit SIC code.

This large dataset allows us to construct more accurate measures of innovation and competition and to test the robustness of our findings when different outcomes are used to capture technological change and productivity growth. Specifically, innovation is measured with two different sets of variables. The first set consists of number of patents and number of citations received by those patents. As opposed to a simple patent count, citations can capture not only the quantity of ideas produced but also the quality of those ideas. The main advantages of patents compared to other R&D indicators are that they provide a measure of successful research output and they are objective in so far as they are not influenced by accounting practices. At the same time, there are a number of limitations in measuring innovation through patents, which we have already discussed and documented in Section 3 above. As we showed there, patents and citations measure only a relatively tiny fraction of the actual output of innovative activity. Moreover, patents cannot account for efficiency gains due to the adoption of the most efficient technologies and best managerial practices as long as these are not patented, nor can they account for the increased productivity that follows the introduction of new goods and services not covered by patents.

Coherent with the discussion carried out in Sections 3 and 4, and in order to overcome these limitations and to provide a check of robustness of our findings, the second variable we use to capture technological advances refers to firms' productivity growth computed either as TFP or as labor productivity for 85 different NAICS-4 manufacturing industries over the period 1987–2008, as calculated on the basis of either NBER or BLS output data. Competition is an index based on the profitability of the industry and takes values from 0 (low competition/high profits) to 1 (high competition/low profits).

On the basis of the data so organized, first we have replicated and extended the empirical model discussed in Section 3 (see Correa 2010), with innovation (patents or patent citations) on the left-hand side and competition (inverse of profitability) on the right-hand side. In line with the theoretical arguments developed in Section 4 as well as with the empirical results discussed in Section 3, we find a positive relationship that is remarkably robust to changes in industry classification (SIC-4 vs. NAICS-4), time sample period (1990–2001 vs. 1975–2001), and set of sampled industries (manufacturing vs. all industries).

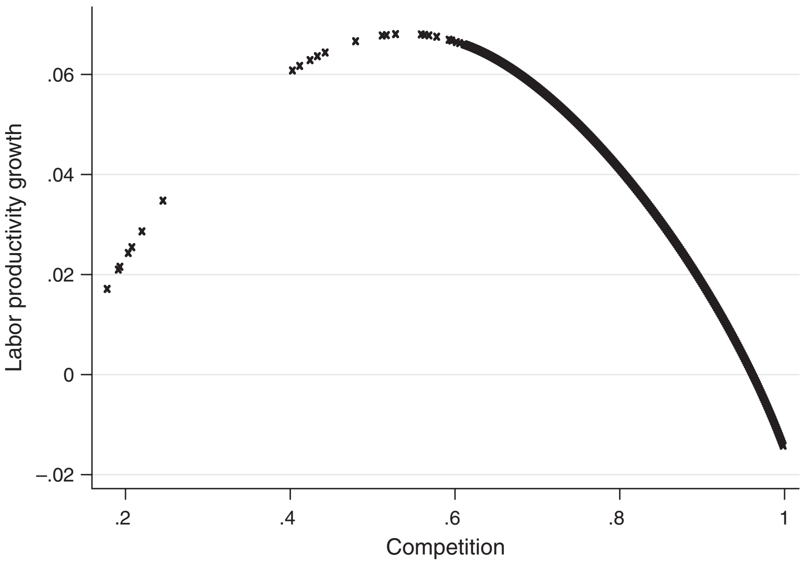
Granted, then, that the traditional measures of innovations are—in our dataset very much like those considered in Section 3—positively correlated with the most natural index of competitive pressure, we moved next to the issue that is most important for us, that is, the correlation between the latter and objective measures of productivity growth. The next two figures (Figures 7 and 8) are based on NAICS-4 industries for the period 1987–2007. Figure 7 seems to suggest that, when one regresses productivity growth on the inverse of profitability, the inverted-U relationship that had been pushed out the door in the case of patents strikes back with a vengeance when studying labor productivity growth.

Nevertheless, as amply discussed in Correa and Ornaghi (2011) and Hashmi (2011), the (inverse of) profitability is clearly an endogenous outcome that depends, among other things, on the ability of firms (in each sector) to reduce costs and increase prices through innovations that increase labor productivity.

To deal with this problem, Figure 8 is obtained by regressing labor productivity growth on lagged inverse profitability, that is, on measured sectorial competition during the previous period. This ''simple'' change has a dramatic effect: from an inverted-U we move to a clear positive relationship—and a very robust one as the tests reported in Correa and Ornaghi (2011) show.

This result is confirmed when the observations are aggregated sector by sector and the regression is performed using the sectorial

Figure 7
Labor Productivity Growth and Competition



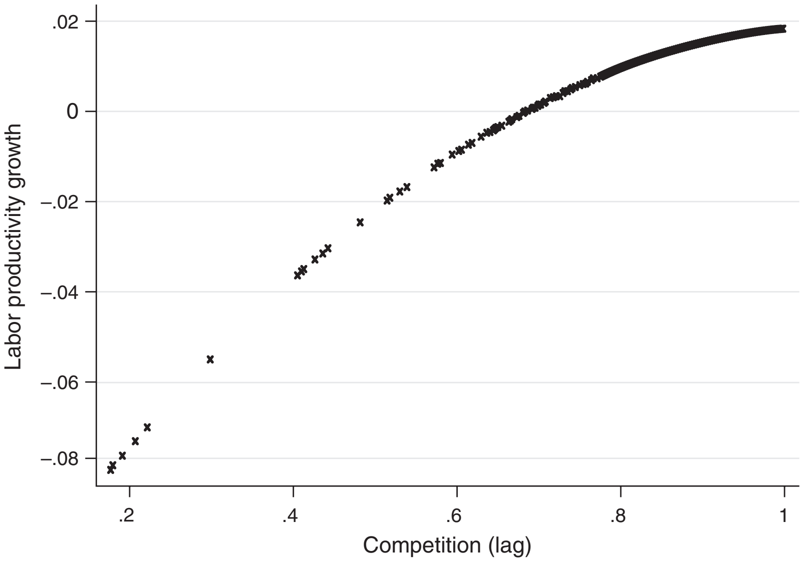
averages on both sides of the equality sign. If possible, the positive relationship displayed in Figure 9 is more pronounced than that in Figure 8.

It is worth stressing that, because of re-scaling, the interval of variation of productivity growth, as reported in Figure 9, may incorrectly appear as exceedingly small. This is not the case; when expressed in percentage points per year, the effect of increased competition is actually quite strong. Once the few very monopolized sectors in which our index of competition is lower than 0.7 are dropped, the actual lower bound is a growth rate of 1.5 percent per year (recall that these are sectorial averages), and the upper bound is 3.5 percent. In other words, the average annual growth of productivity in the sectors with the highest level of competition is up to 2 percent bigger than in the sectors with the lowest level of competition. These are strikingly large differences when cumulated over various decades, as is the case in our dataset.

6. OPEN ISSUES AND CONCLUSIONS

Economic theory is ambiguous when it comes to assessing if either competition and free entry or patent protection and the monopoly

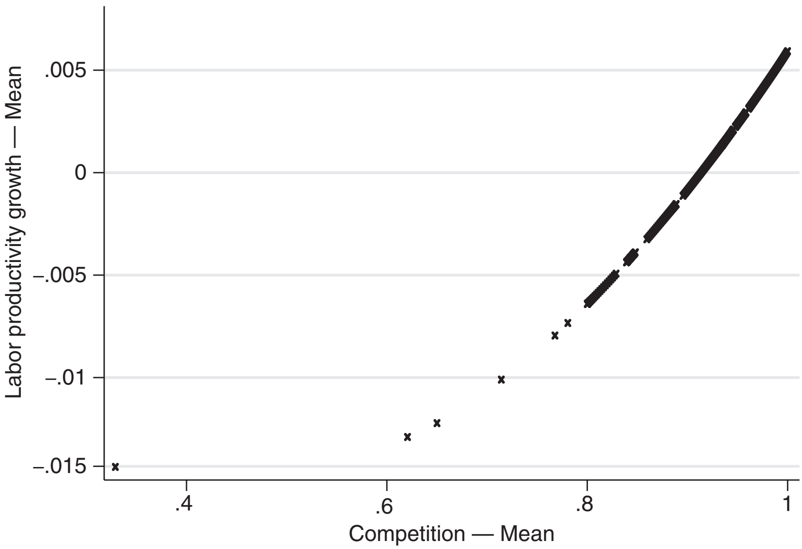
Figure 8
Labor Productivity Growth and Lagged Competition



power it induces define the best institutional environment to foster innovation and technological progress. The empirical evidence, though, is overwhelming: far from fostering and stimulating innovation and productivity growth, patents are likely to hamper them, while free entry and competition appear to be at least correlated with (if not the cause of) labor productivity and TFP growth. This has an obvious implication: There are no objective reasons whatsoever to strengthen patents any further than we have already done. In fact, it seems urgent to begin undoing some of the damage we have been doing to ourselves over the last 30 years and work to reform—slowly but surely—the overall system of intellectual property.

Reforming the system is neither politically easy, nor is it obvious, at least to us, which steps are appropriate and in which order they should be undertaken. As a matter of fact, the number of empirical studies that have carried out well-designed counterfactual exercises capable of assessing how an alternative legal system would affect innovative activity and which parts of the current one are the most damaging to technological progress is exceedingly small. What this

Figure 9
Labor Productivity Growth and Lagged Competition
(Sectorial averages)



means is that the very first policy goal may just be that of advocating for, supporting, and carrying out well-designed empirical studies of how innovation and technological progress fare across sectors, countries, and time periods, under different legal and incentive systems. Our basic intuition, grounded on the research either presented or referred to here, is that the current systems of intellectual property are tantamount to the trade restrictions in existence until a few decades ago, and their dismantlement should be approached in the very same fashion, piece by piece and quite patiently.

We do not have, therefore, any grand strategy to propose, just a short list of general goals plus a few somewhat more specific ones in those particular areas where, we believe, economic analysis has managed to dig deeper. Because policy proposals are better digested and metabolized when served in the form of small pills, here is our list:

- Stop the still-rising tide that, since the early 1980s, is both extending the set of ''things'' that can be patented and shifting the legal and judicial balance more and more in favor of patent holders.

- Because competition fosters productivity growth, antitrust and competition policies should be seen as key tools to foster innovation. This is of particular relevance for high-tech sectors, from software, to bioengineering, to medical products and pharmaceuticals.
- Free trade is a key part of competition policies, hence the role that the World Trade Organization, World Intellectual Property Organization, and Trade Related Aspects of Intellectual Property Rights agreement play should be redefined to move away from the current neomercantilist approach to free trade in goods and ideas. The aim here should be that of stopping the policy of exporting our intellectual policy laws to other countries while adopting a policy of exporting free trade and competition in innovation. This seems to us an urgent goal because, within a couple of decades, the ''balance of trade in ideas'' between the United States and Europe on one hand and Asia on the other may easily reverse. At that point, the temptation to engage in ''mercantilism of ideas'' may well affect the now-developing Asian countries, leading to a general increase in intellectual property protection worldwide.
- Cross-industry variation in the importance of patents suggests we may want to start tailoring patent length and breadth to different sectorial needs. Substantial empirical work needs to be done to implement this properly, even if there already exists a vast legal literature pointing in this direction.
- Reverse the burden of proof. Patents should be allowed only when monopoly power is justified by evidence about fixed costs and actual lack of appropriability. The operational model should be that of ''regulated utilities'': patents should be awarded only when strictly needed on economic grounds.
- Use prizes and competition to nurture innovation. An interesting approach is to change the role that the National Science Foundation and the National Institutes of Health play in fostering innovation. The basic goal, in this case, is that of reversing the principle according to which federally financed investigation can lead to private patents. As a first step, we would advocate going back to the old rule according to which the

results of federally subsidized research cannot lead to the creation of new private monopolies but should be available to all market participants. This goal is particularly important in the pharmaceutical industry.

- With regard to the pharmaceutical industry, we advocate reforming pharmaceutical regulation to either treat Stage II and III clinical trials as public goods (to be financed by NIH on a competitive basis) or by allowing the commercialization (at regulated prices equal to the economic costs) of drugs that satisfy the Food and Drug Administration requirements for safety even if they do not yet satisfy the current, over-demanding requisites for proving efficacy. It is ensuring the efficacy—not the safety—of drugs that is most expensive, time-consuming, and difficult. All the usual mechanisms of ensuring the safety of drugs would remain firmly in place. While pharmaceutical companies would be requested to sell new drugs at ''economic cost'' until efficacy is proved, they could start selling at market prices after that. In this way, companies would face strong incentives to conduct or fund appropriate efficacy studies where they deem the potential market for such drugs to be large enough to bear the additional costs. At the same time, this ''progressive'' approval system would give cures for rare diseases the fighting chance they currently do not have. This solution would substantially reduce the risks and cost of developing new drugs.
- If this progressive approval approach works for rare diseases, there is no reason it should not be adopted across the board. The current system favors a small number of blockbuster drugs that can be sold to millions of patients. The coming revolution in medicine will rely on carefully targeting hundreds or even thousands of drugs to the correct patients. But lawmakers must first usher in a new system that makes developing these precision treatments possible. The regulation reform we are suggesting here would be a first important step to achieve such a goal.
- Finally, software patents are a particularly egregious and bad form of intellectual property for a sector where we also have very detailed micro evidence about the role of patents in (not) promoting innovation (see, e.g., Bessen and Meurer

2008 and references therein). The same arguments are likely to apply to bio-engineering and genetic research at large. The goal of policy, in these cases, should be just that of slowly but surely decreasing the strength of intellectual property interventions.

7. APPENDIX

Figures A1 and A2 reproduce the above Figures 1–2 and 3–4, respectively, with each industry's NAICS-4 number appended to its data point. This should allow the interested reader to make up his mind about the extent to which, sector by sector, patent citations counts are, or are not, a good proxy for those factors that, in practice, do increase productivity. We remind readers that these are averages over the 20-year period 1987–2007.

Figures A3 and A4 are from the same dataset. They have labor and TFP growth on the vertical axes, respectively, and citations growth on the horizontal axes. There is no correlation in either figure.

Figures A5–A8 should allow for a better understanding of the findings reported in Section 5. A5 and A6 show how the inverted-U relationship and the positive relationship fit the actual observations. Both curves are, overall, quite flat. A7 and A8 replicate those displayed in Section 5. They are a ''zoom'' of the fitted line reported above. The graphs reveal the ''fragility'' of the empirical analysis based on these specifications and underline the need for additional research.

Figure A1
Patent Citations and Labor Productivity Growth
(Right panel excludes the top 5 percent of performers)

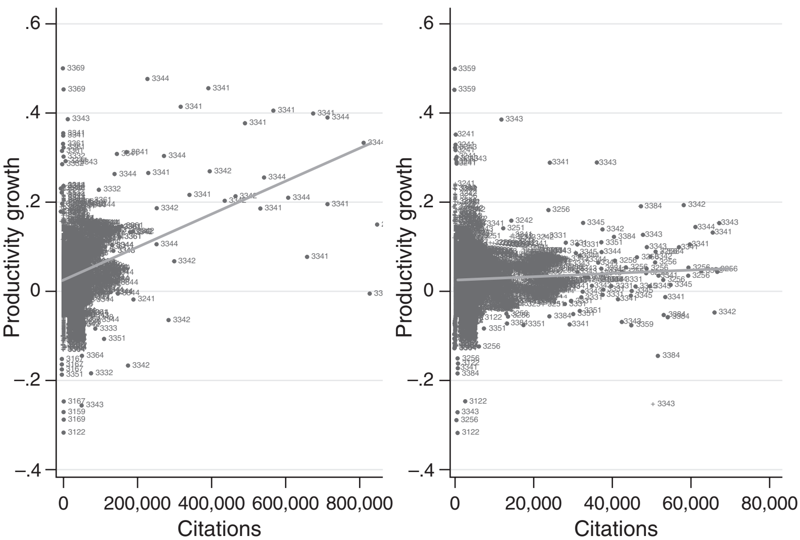


Figure A2
Average Citations and Average Labor Productivity Growth
(Right panel excludes the top 5 percent of performers)

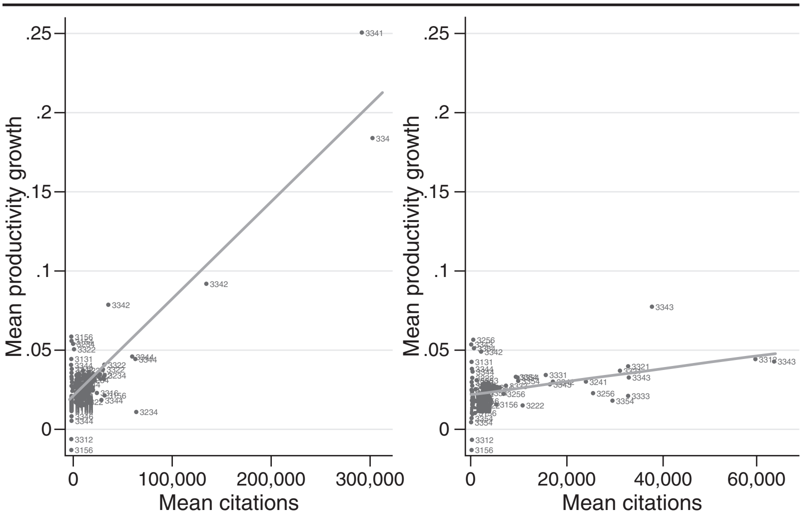


Figure A3
Citation Growth and TFP Growth

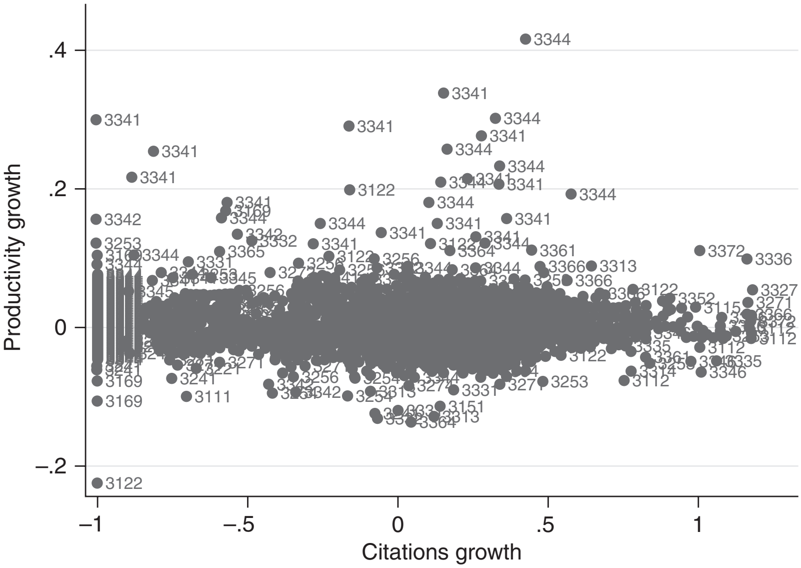


Figure A4
Citation Growth and Labor Productivity Growth

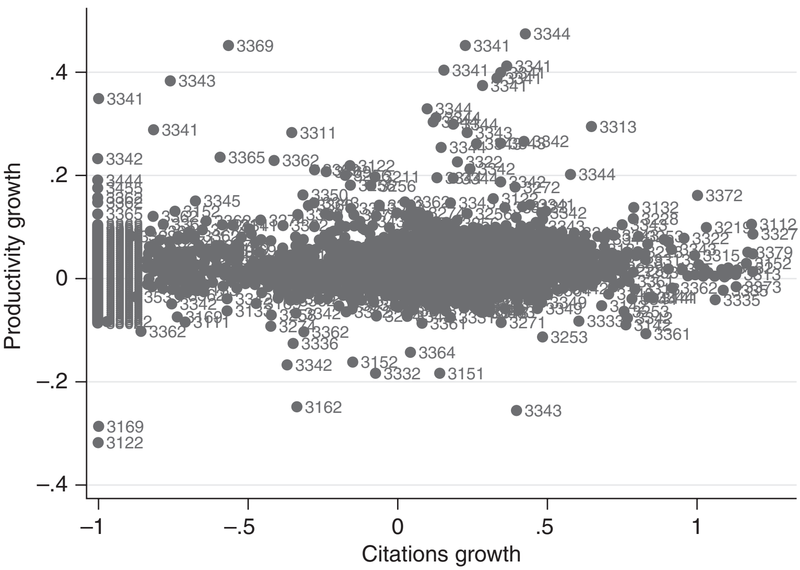


Figure A5
Relationship between Competition and Labor Productivity Growth

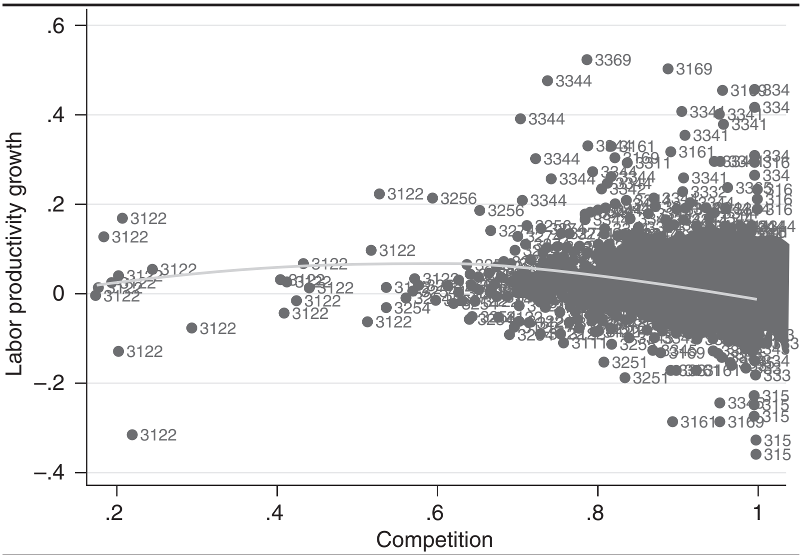


Figure A6
Relationship between Lagged Competition and Labor Productivity Growth

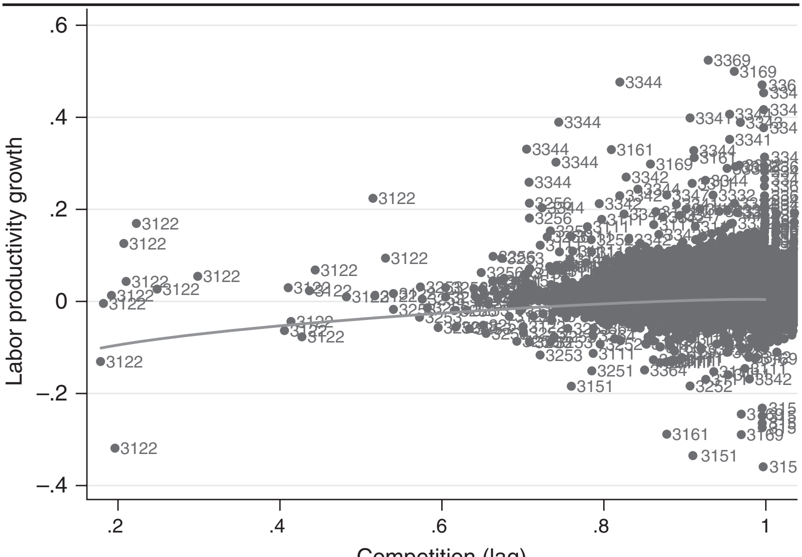


Figure A7
Labor Productivity Growth and Competition

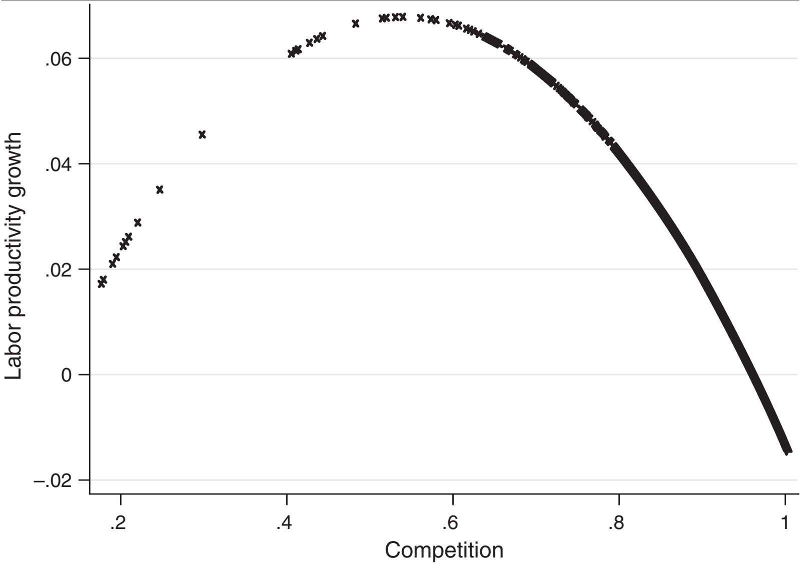
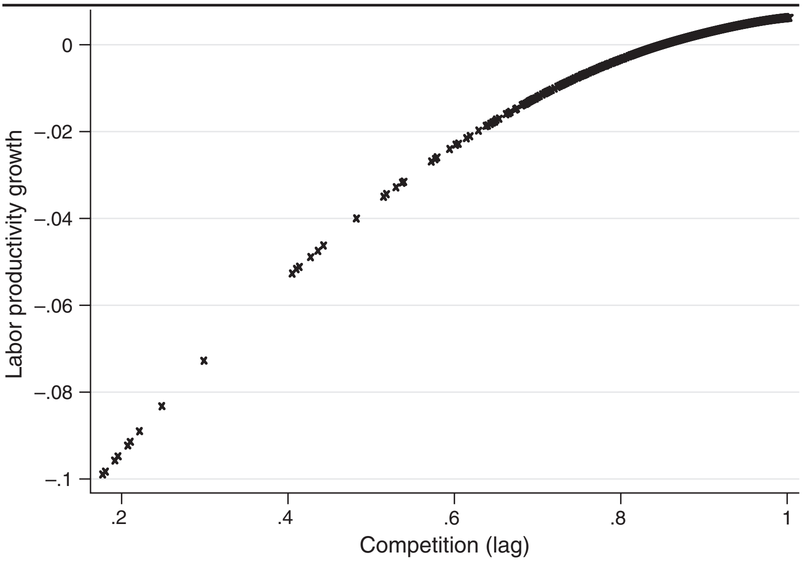


Figure A8
Labor Productivity Growth and Lagged Competition



REFERENCES

Adams, J., and A. Jaffe. 1996. ''Bounding the Effects of R & D: An Investigation Using Matched Establishment–Firm Data.'' *RAND Journal of Economics* 27 (4): 700–21.

Aghion P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2005. ''Competition and Innovation: An Inverted U Relationship.'' *Quarterly Journal of Economics* 120: 701–28.

Aghion, P., and T. Griffith. 2005. *Competition and Growth: Reconciling Theory and Evidence*. Cambridge, MA: MIT Press.

Aghion, P., and P. Howitt. 1992. ''A Model of Growth through Creative Destruction.'' *Econometrica* 60 (2): 323–51.

Arrow, K. 1962. ''Economic Welfare and the Allocation of Resources for Invention.'' In *The Rate and Direction of Innovative Activity*, ed. R. Nelson. Princeton, NJ: Princeton University Press.

Bartelsman, E. J., and W. Gray. 1996. ''The NBER Manufacturing Productivity Database.'' National Bureau of Economic Research Technical Working Paper No. 205.

Baumol, W. J. 2010. *The Microtheory of Innovative Entrepreneurship*. Princeton, NJ: Princeton University Press.

Baumol, W. J., R. E. Litan, and C. J. Schramm. 2007. *Good Capitalism, Bad Capitalism, and the Economics of Growth and Prosperity*. New Haven, CT: Yale University Press.

Bessen, J., and R. M. Hunt. 2003. ''An Empirical Look at Software Patents.'' Mimeo. Abridged version in *Journal of Economics and Management Strategy* 16: 157–89.

Bessen, J., and M. J. Meurer. 2008. ''Of Patents and Property.'' *Regulation* 32 (4): 18–26.

Blundell, R., R. Griffith, and J. Van Reenen. 1999. ''Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms.'' *Review of Economic Studies* 66: 529–54.

Boldrin, M., and D. K. Levine. 1997. ''Competitive Equilibrium Growth.'' Mimeo. Universidad Carlos III and University of California, Los Angeles. October.

———. 2004. ''Rent Seeking and Innovation.'' *Journal of Monetary Economics* 51: 127–60.

———. 2005. ''The Economics of Ideas and Intellectual Property.'' *Proceedings of the National Academy of Sciences* 102: 1252–6.

———. 2006. ''Quality Ladder, Competition and Endogenous Growth.'' Mimeo. Washington University in St Louis.

———. 2008a. *Against Intellectual Monopoly*. Cambridge, UK: Cambridge University Press.

———. 2008b. ''Perfectly Competitive Innovation.'' *Journal of Monetary Economics* 55: 435–53.

———. 2009a. ''IP and Market Size.'' *International Economic Review* 50: 855–81.

———. 2009b. ''A Model of Discovery.'' *American Economic Review: Papers and Proceedings* 99: 337–42.

Carlin, W., M. Schaffer, and P. Seabright. 2004. ''A Minimum of Rivalry: Evidence from Transition Economies on the Importance of Competition for Innovation and Growth.'' *Berkeley Electronic Journal of Economic Analysis and Policy: Contributions* 3: 1–43.

Castiglionesi, F., and C. Ornaghi. 2011. ''On the Determinants of TFP Growth: Evidence from Spanish Manufacturing Firms.'' Mimeo. Tilburg University and University of Southampton, 2008. Forthcoming in *Macroeconomic Dynamics*.

Cohen, W. M., R. R. Nelson, and J. P. Walsh. 2000. ''Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not).'' NBER Working Paper No. 7552. February.

Correa, J. A. 2010. ''Innovation and Competition Relationship in a Memory Process.'' Ph.D. dissertation. UK: University of Southampton. Forthcoming in *Journal of Applied Econometrics* under the title ''Innovation and Competition: An Unstable Relationship.''

Correa, J., and Ornaghi, C. 2011. ''Competition and Innovation: New Evidence from U.S. Patent and Productivity Data.'' Unpublished manuscript. UK: University of Southampton.

Dean, E. R., and M. J. Harper. 1998. ''The BLS Productivity Measurement Program.'' U.S. Bureau of Labor Statistics.

Gilbert, R., and C. Shapiro. 1990. ''Optimal Patent Length and Breadth.'' *Rand Journal of Economics* 21: 106–12.

Griliches, Z. 1957. ''Hybrid Corn: An Exploration of the Economics of Technological Change.'' Ph.D. dissertation. Chicago: University of Chicago.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg. 2001. ''The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools.'' NBER Working Paper No. 8498.

Hashmi, A. R. 2011. ''Competition and Innovation: The Inverted-U Relationship Revisited.'' Mimeo. National University of Singapore. February.

Hellwig, M., and A. Irmen. 2001. ''Endogenous Technical Change in a Competitive Economy.'' *Journal of Economic Theory* 101: 1–39.

Jaffe, A. B., and J. Lerner. 2004. *Innovation and Its Discontents*. Princeton, NJ: Princeton University Press.

Klette, T. J., and S. Kortum. 2004. ''Innovating Firms and Aggregate Innovation.'' *Journal of Political Economy* 112: 986–1018.

Kortum, S., and J. Lerner. 1998. ''Stronger Protection or Technological Revolution: What Is Behind the Recent Surge in Patenting?'' *Carnegie-Rochester Conference Series on Public Policy* 48: 247–304.

Legros, P. 2005. ''Art and Internet: Blessing the Curse?'' Mimeo. Belgium: ECARES and the Université Libre de Bruxelles.

Lerner, J. 2009. ''The Empirical Impact of Intellectual Property Rights on Innovation: Puzzles and Clues.'' *American Economic Review: Papers and Proceedings* 99 (2): 343–8.

Levin, R. C., A. K. Klevorick, R. R. Nelson, and S. G. Winter. 1987. ''Appropriating the Returns from Industrial Research and Development.'' *Brookings Papers on Economic Activity* 3: 783–820.

Liebowitz, S. J. 1985. ''Copying and Indirect Appropriability: Photocopying of Journals.'' *Journal of Political Economy* 93: 945–57.

Moser, P. 2005. ''How Do Patent Laws Influence Innovation?'' *American Economic Review* 95 (4): 1214–36.

Nickell, S. J. 1996. ''Competition and Corporate Performance.'' *Journal of Political Economy* 104: 724–46.

Okada, Y. 2005. ''Competition and Productivity in Japanese Manufacturing Industries.'' NBER Working Paper No. 11540.

Pakes, A. S. 1986. ''Patents as Options: Some Estimates of the Value of Holding European Patent Stocks.'' *Econometrica* 54: 755–84.

Piazza, R. 2010. ''Leadership Contestability, Monopolistic Rents and Growth.'' Mimeo. International Monetary Fund. December.

Quah, D. 2002. ''24/7 Competitive Innovation.'' Mimeo. London School of Economics.

Scherer, F. M. 1990. *Industrial Market Structure and Economic Performance*. Boston: Houghton Mifflin.

Schumpeter, J. 1942. *Capitalism, Socialism and Democracy*. New York: Harper and Brothers.

Scotchmer, S. 1991. ''Standing on the Shoulders of Giants: Cumulative Research and the Patent Law.'' *Journal of Economic Perspectives* 5: 29–41.

Sheshinski, E., R. J. Strom, and W. J. Baumol, eds. 2007. *Entrepreneurship, Innovation, and the Growth Mechanism of the Free-Enterprise Economies*. Princeton, NJ: Princeton University Press.

Solow, R. M. 1956. ''A Contribution to the Theory of Economic Growth.'' *Quarterly Journal of Economics* 70 (1): 65–94.

———. 1957. ''Technical Change and the Aggregate Production Function.'' *Review of Economics and Statistics* 3 (3): 312–20.

Stigler, G. J. 1956. ''Industrial Organization and Economic Progress.'' In *The State of the Social Science*, ed. L. D. White, pp. 269–82. Chicago: University of Chicago Press.

Van Reenen, J. 2010. ''Does Competition Raise Productivity through Improving Management Quality?'' CEP Discussion Paper No. 1036. December.

Comment

Samuel Kortum

As someone who got the ''received wisdom'' about the economics of innovation as a part of his schooling, I approached the ''economic heresy'' of Michele Boldrin and David Levine with some skepticism. I was also intrigued. This paper, with coauthors Juan Correa and Carmine Ornaghi, (hereafter BCLO) discusses those theoretical arguments, presents empirical evidence, and draws out the policy implications. While I don't think they make a convincing case for dismantling the patent system, I do find parts of the argument compelling.

COMPETING THEORIES OF TECHNOLOGICAL CHANGE

In the introduction to his 1990 article in the *Journal of Political Economy*, Paul Romer articulates what I think of as the received wisdom. It starts with a careful account of the production function. Economic output is produced with factors of production, unskilled labor, skilled labor, machinery, etc. Technology is the recipe book for how to combine these ingredients (the factors of production) to produce goods that we like. In principle, you could always make twice as much of any good you're currently making by using twice the quantity of each ingredient. You don't need a new recipe to do that. Similarly, with the same recipe, anyone else could produce more of what you are currently making. A recipe is said to be *nonrival*, as any number of people can use it simultaneously. That's why, in a competitive equilibrium, all the revenue from production is paid to the factors of production, with nothing left to pay for the recipes.

Numerous studies have shown that the growth of factors of production leaves much to be explained in accounting for economic growth. It seems that the recipes must be getting better, and casual

Samuel Kortum is professor of economics at the University of Chicago.

observation, for example using a new smart phone, supports that view. Since technological change appears to be a crucial source of rising living standards, yet a competitive market doesn't provide incentives for inventive activity, we are fortunate to have institutions that fill the gap. The patent system is one of them, allowing the owner of a patented recipe to restrict its use by others. The patent holder gets some monopoly power. The expectation of monopoly profit is an incentive to invent new recipes.

What could be wrong with this received wisdom? The starting point of the ''economic heresy'' is the observation that it can be difficult to separate the recipe from the chef. New techniques are often embodied in people or machines. It takes time and effort to train people or build machines. While a component of the technique may be nonrival in principle, it may not be easy to replicate in practice. The good techniques end up being used by some and not by others. A consequence is that there can be rents to the inventor of a new technique, even in the absence of patent protection.

This starting point is quite compelling and is, in fact, a common view of technology. A recent paper by Robert Lucas (2008), while not modeling any payments, envisions the diffusion of techniques in this way. A technique only spreads from one individual to another via a chance meeting and an ensuing transfer of information (from one brain to the other). Note that techniques are nonrival goods in this economy. But, because they are embodied in individuals, the good techniques aren't instantaneously replicated. In this economy, one could imagine that an individual with knowledge of a good technique might put a price on the service of teaching it to someone else.

These arguments become less abstract with a specific example in mind. Consider a new technique for folding metal into various shapes without weakening it. The idea behind it is nonrival. A good engineer could look at it, get the idea, and decide how it could be put to use in his manufacturing process. Without patent protection, the inventor of this technique would seem to have no means of extracting any return from it. But actually adopting the new process of folding metal requires educating and training engineers. The inventor knows how to make it work, but the potential users do not. In principle, even in the absence of a patent, the inventor could profit from manufacturers who hire him as a consultant to get the process up and running.

THE CHALLENGE FOR EMPIRICAL WORK

Faced with these two scenarios and their starkly different policy implications for the patent system, how can we use data to resolve the issue? In principle, we'd like to build an economic model that encompasses these two scenarios. The model would also incorporate features of the current patent system, as well as other mechanisms inventors use to appropriate returns, such as trade secrecy. We would then confront this model with microeconomic data of various sorts in order to estimate the relevant parameters. Finally, using these estimates of the parameters, we would perform various counterfactual simulations of the model to examine the outcomes for economic growth and economic welfare under various possible reforms of the patent system. We could then make policy recommendations with some confidence, based on such a systematic quantitative analysis.

Back to reality: While a worthy goal, we have a long way to go before we can carry out such a research agenda in a satisfactory manner. In the meantime, we're left looking at correlations between measures of productivity, patents, firm size, R&D, and competition (measured as the inverse of profitability). We have to be pretty cautious about what we can conclude from such correlations about the worthiness of either scenario. The implications for patent policy are even more tenuous.

In some cases, the mapping from the empirical findings to the different models of innovation is not very clear. The received wisdom is not that large firms are the most inventive. As formalized in economic models, all the innovation is done by new entrants. Similarly, since the models associated with the received wisdom typically assume free entry into research activity, it is not clear that the industries with more research activity will appear less competitive (more profitable). In these models, profits get used up paying for research.

The paper is quite right, on the other hand, to point to the weak connection between productivity growth and patenting. It is a genuine puzzle on many levels that the industries with more inventive activity do not experience much more rapid productivity growth. Furthermore, this puzzle persists if you replace patents with R&D spending. In his presidential address to the American Economic Association, Zvi Griliches (1994) dwelled on this issue. Figure 2 in his paper gives roughly the same message as Figures 1–4 in BCLO.

Why are these relationships so weak? Griliches pointed to the difficulty of measuring productivity. There are also conceptual issues. An industry experiencing technological improvements will typically expand into less productive activities, so that the expansion itself may dampen measured productivity growth. This point is carefully worked out in a recent paper by Costinot, Donaldson, and Komunjer (2011).

The empirical work in BCLO is a useful reminder that we ought to be humble given how little we know in this field. The authors suggest that the weakness of the evidence in favor of the received wisdom means that there is little justification for our current patent system. Given our limited knowledge, I think the appropriate response is to be very cautious in recommending dramatic changes in that institution.

OTHER PROMISING APPROACHES

My critique of the empirical work in the paper may appear unconstructive. What *is* a good way of determining which view of innovation and competition is more reasonable? Short of the rather heroic approach I laid out in the beginning, how else might we use data to answer the important question of how we can best reform the patent system? I see at least three useful lines of attack:

- examining the key premises of the economic heresy view,
- exploiting macro evidence on patent-system reform, and
- conducting case studies on how actual inventors and users of new technology interact with the patent system.

I consider each of these in turn.

Key Premises

The economic heresy view is built on the premise that technology, while nonrival, is not easy to replicate. That premise appears very solid but ripe for empirical exploration. At the macro level, it relates to why many countries are not exploiting state-of-the-art technology. At the micro level, it relates to why individual technologies don't spread faster. For example, even Wal-Mart took a long time to spread across the United States, as graphically illustrated in Holmes (2011). But what is it that holds things back? Is it that technology is inherently costly to replicate, generating rents to inventors? Or is it that users, either consumers or producers, are resistant to new technology, thus reducing the discounted flow of profits to inventors? A better understanding of this phenomenon of slow replication seems central to the issues of this paper and economics more generally.

Macro Evidence

Many countries have recently strengthened patent enforcement. Sometimes these changes have been driven by internal pressure from their own inventors and sometimes by external pressure from developing countries who want their intellectual property protected. In either case, these changes provide a laboratory within which to study the consequences of changing the patent system. While Frederic Scherer is mentioned in BCLO as buying into the received wisdom, I have seen him point to evidence that when Italy adopted stronger pharmaceutical patents, its pharmaceutical industry suffered.

Case Studies

Finally, in this area, case studies can be extremely valuable. In preparing for this discussion, I had a long conversation with my cousin Max W. Durney, the inventor of the technology for folding metal that I mentioned above. On the basis of this invention, which is protected with numerous patents, he founded the company Industrial Origami. He argues that his extensive use of patents was crucial in attracting financing. Up-front financing was itself necessary to invest in developing the technology and in marketing it to manufacturers who were initially resistant to adopting a new process. The current patent system, while not without problems, can enable creative individuals to make a career inventing new technologies.

REFERENCES

Costinot, A., D. Donaldson, and I. Komunjer. 2011. ''What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas.'' *Review of Economic Studies*. Forthcoming.

Griliches, Z. 1994. ''Productivity, R&D, and the Data Constraint.'' *American Economic Review* 84: 1–23.

Holmes, T. J. 2011. ''The Diffusion of Wal-Mart and Economies of Density.'' *Econometrica* 79: 253–302.

Lucas, R. E. 2008. ''Ideas and Growth.'' *Economica* 76: 1–19.

Romer, P. M. 1990. ''Endogenous Technical Change.'' *Journal of Political Economy* 98 (5): 71–102.

Comment

Andrew Atkeson

Michele Boldrin and David Levine, in previous work and now in this paper with Juan Correa and Carmine Ornaghi, have issued a powerful challenge to the idea that patent protection plays an essential role in fostering technological innovation and improved welfare for consumers. In fact, these authors go further at times and argue that patent protection may in fact hinder rather than help technological innovation.

The authors start this paper with the premise that it is a straightforward task to construct economic models such that patent protection improves welfare when the models' parameters are in one region and reduces welfare when the models' parameters are in another region. They argue that, as a result, the question of whether patent policy has a beneficial or harmful effect on innovation and consumer welfare must be resolved on the basis of empirical evidence rather than theory. In this paper, the authors aim to shed some light on the empirical link between patents, competition, and technological progress.

I anticipate that in his discussion, commenter Sam Kortum will assess the contribution of this paper to the broader empirical literature on the relationship between patenting, innovation, and technological progress. Given the dictates of comparative advantage, I will specialize in my discussion on the theoretical basis for arguing that patent protection may be harmful rather than helpful to innovation and welfare. I agree with the basic premise of the paper's authors that the idea that patents are essential to support innovation in equilibrium is deeply ingrained in many economists' and policymakers' minds. While many practitioners would argue that there are important flaws in patent policy as currently implemented in the

---

Andrew Atkeson is the Stanley M. Zimmerman Professor of Economics and Finance at the University of California, Los Angeles.

United States (see, e.g., Shapiro 2007), the authors of this paper look to go further and argue that the property rights conferred by patent protection, even if executed well, may be harmful to innovation and welfare. This view is the ''economic heresy'' the authors refer to in their introduction. Given the strength of the authors' theoretical arguments, I see my best chance of adding value to this debate is to sketch a simple theoretical model of innovation with and without patent protection to see how straightforward it is to have the welfare implications of patents go either way.

My goal in sketching this model is to capture some of the ideas the authors discuss in their paper about the difference between *innovation* and *imitation*, and the role of a *first-mover advantage* for innovators of some kind in supporting innovation without patent protection in a way that complements the previous theoretical work by these authors. The authors are certainly correct that the theoretical work assessing the impact of patent policy in general equilibrium in the ''New Growth Theory'' literature is too special and narrow and that much more theoretical and empirical work needs to be done. What follows is a small contribution in that direction.

The central idea I would like to illustrate with this model is that, in an economy with imperfect competition, incumbent firms built on a successful innovation already have strategies available to preempt entry by imitating firms even in the absence of patent protection. As a theoretical matter, it is not at all clear that welfare is improved if we enhance the strategic position of these incumbents by granting them patent protection as an additional tool to deter entry of competing firms.

The model I use to illustrate this point starts with a demand structure in which we can say that some products are closer substitutes than others, so that we can think about oligopoly in a particular market nested in a larger general equilibrium economy as a whole. To do so, posit a nested demand system with a continuum of potential *sectors* and many (but countable) potential *goods* within each sector. A firm *innovates* by being the first to introduce a good into a particular sector. To innovate, a firm must pay a high fixed cost to introduce this first good in the sector. A firm *imitates* by being the second or later to introduce a good into a particular sector. An imitator benefits from following an innovator into a sector in that the imitator pays a lower fixed cost to introduce this additional good in the sector.

To be more specific, let aggregate consumption be a constant elasticity of substitution aggregate (CES) across sectors with

$$C = \left[\int_0^N y_j^{1-\frac{1}{\eta}} \, dj \right]^{\frac{\eta}{\eta-1}}$$

Here y_j is the output of sector j and N is the measure of sectors that have active firms producing goods in those sectors. The parameter η is the elasticity of substitution across sector inputs in producing final consumption. Standard arguments give that the induced demand by final consumption producers for the output of sector j is given by the CES (inverse) demand function

$$\frac{P_j}{P} = \left(\frac{y_j}{C} \right)^{-\frac{1}{\eta}}$$

where P_j is the price index for output of sector j and P is the price index for final consumption C.

Output in sector j is a second CES aggregate across the output of the K_j firms active in that sector

$$y_j = \left[\sum_{k=1}^{K_j} q_{j_k}^{1-\frac{1}{\rho}} \right]^{\frac{\rho}{\rho-1}}$$

where q_{jk} is the output of the kth firm in sector j and ρ is the elasticity of substitution across goods within the sector. To capture the idea that goods within a sector are closer substitutes than goods in different sectors, assume that $\rho > \eta$. With this ranking of elasticities, we have the inverse demand curve for a particular good given by

$$\frac{P_{jk}}{P_j} = \left(\frac{q_{jk}}{y_j} \right)^{-\frac{1}{\rho}}$$

where P_{jk} is the price of the kth good in sector j and the price index for the sector P_k is constructed in the standard manner.

Now consider some simple economics of firms' decisions to innovate and imitate when facing this demand structure. The returns to either of these decisions will depend on the number of firms that introduce goods in a particular sector. Assume that when it comes time to produce, firms active in a sector engage in price (Bertrand) competition. With a finite number of goods in a sector, the perceived elasticity of residual demand for each firm in a sector—and hence the markup over marginal cost charged by that firm—depends on that firm's share of the market in that sector.[1] If there are no imitators that follow an innovating firm into a particular sector, then that innovating firm enjoys a high profit from its innovation not only because it commands the entire market but also because it charges a high markup corresponding to the low elasticity of substitution η across sectors. If many imitators follow an innovating firm into a sector, each firm has a small market share in the sector, and markups are small, corresponding to the high perceived elasticity of residual demand for each firm as determined by ρ. A firm that innovates pays an entry cost c_1, while a firm that imitates pays an entry costs c_2, with $c_1 \geq c_2$.

Are patents *necessary* to support innovation in this environment? The answer to this question depends on parameters.

If the cost of imitation is zero and there are potential profits for an imitator (i.e., $\rho < \infty$), then the answer is *yes*. We cannot have an equilibrium in which an incumbent innovator earns the positive profits necessary to recoup the innovation cost $c_1 > 0$ without attracting entry from imitators. This particular parameter configuration corresponds to the standard assumption that imitation is truly costless, and in this special case, it is likely impossible to support innovation in equilibrium without some regulation of imitation such as patent protection. As the authors of this paper correctly point out, this special case of zero imitation costs likely lies at the heart of most thinking by policymakers about intellectual property.

How general is this argument for patent protection? Not very. Consider, for example, the case in which goods within a sector are perfect substitutes ($\rho = \infty$) and imitation costs are positive ($c_2 > 0$). In this case, with goods within a sector being perfect substitutes,

[1] This dependence of elasticities on market shares arises as a result of the assumption that there are only a finite number of goods K_j being produced in any given sector. As a result, firms must take into account that the price that they choose affects the sectoral price index P_j and thus overall demand for the sectoral output y_j in addition to the share of spending in the sector on their specific good. In contrast, with a continuum of sectors, individual firms take the overall consumption price index P and level of consumption C as given.

entry by a second firm as an imitator in a sector eliminates all profits for both the innovator and the imitator by driving prices down to marginal cost through Bertrand competition. Here, there are no returns to imitation. All firms prefer to innovate in a new sector and enjoy monopoly profits in that sector rather than imitate. In this extreme case, innovators here are protected from imitation as long as there is some cost to imitation.[2] Similar logic holds that patent protection is also unnecessary if imitation costs are equal to innovation costs $c_2 = c_1$ regardless of the elasticity ρ between goods in a sector. In these cases, adding patent protection to the economy has no impact on equilibrium innovation and welfare.

In these special cases, the impact of patent protection on innovation and welfare is easy to see. What happens, however, for more general values of imitation costs relative to innovation costs ($0 < c_2 < c_1$) and elasticities across sectors and goods ($\eta < \rho < \infty$)? Are the impacts of patent policy on innovation, imitation, and consumer welfare purely a problem of parameter values in this more general case? Or are there some more general lessons we might draw even in this simple environment?

It is impossible to give a fully worked-out answer to these questions without specifying the details of the model of entry and the post-entry competition between innovators and imitators, but I conjecture that older ideas about how innovators might preempt entry by imitators in the absence of patent protection may well be useful for drawing more general conclusions about the impact of patents on welfare. The key economic idea to note is that, even in the absence of patent protection, an innovating firm in a particular sector has a *first-mover advantage* in that it can use *product proliferation* through imitation of its own products as a strategy to simultaneously maintain monopoly profits with high markups in equilibrium and to effectively deter entry by imitating firms that might seek to compete with it in the sector.[3]

[2] This idea that firms in a market with imperfect competition will choose to avoid competing with each other in equilibrium is a long-standing one in economics. See, for example, Prescott and Visscher (1977) and Vogel (2008).

[3] See Ellison and Ellison (2011) for recent work documenting how pharmaceutical companies use product proliferation as a strategic device to deter entry by generics as the expiration of their patent protection approaches.

The basic idea is as follows: A single firm that has paid both the innovation cost c_1 for the first good in a sector and the imitation cost c_2 for additional goods in that same sector can charge a high markup for all of those goods by setting high prices (and low quantities) in a coordinated fashion across its product line if there are no other imitators in the sector and, at the same time, credibly convince any potential imitators that the competitive outcome that would emerge if that firm were actually to enter the sector would be very unfavorable to that firm because it would have such a small market share. Essentially, an innovator has the strategic incentive to build on his own innovation through imitation to fill up the sector with his own products. In contrast, if a new firm considers entry into the sector through imitation, it faces the prospect of competition with the original innovator as an incumbent in a powerful position.

What can we conjecture about equilibrium in an economy of this kind in which an innovator can obtain patent protection for his innovation that is broad enough in scope to prevent imitation by other firms looking to introduce goods into the sector? Here patents may well serve the imitation preemption role for innovators that product proliferation serves in the economy without patents. What are the implications of this comparison between an economy without patents and with patents for product creation and welfare? With patents, innovators will clearly charge high prices and enjoy monopoly profits, just as in the equilibrium without patents, so in this respect the two economies might look similar. With patents, however, innovating firms will not feel the competitive pressure from the threat of imitation by other firms to introduce additional products into the sector that they have innovated in, and hence consumers will not benefit from the additional (and relatively cheap) product creation that occurs in the economy without patents when innovating firms fill up their sector with products to deter entry. Hence, by introducing patent protection for innovators, policymakers will in fact have reduced one incentive for product creation without necessarily increasing other incentives to innovate sufficiently to deliver a compensating increase of entry into new sectors.

The overall impact of patent protection on welfare would depend on the general equilibrium reallocation of products across sectors and would require a much fuller analysis of a fully specified model, but the possibilities that patent protection may be a drag on the creation of new products should be clear from the arguments laid out here. What is even clearer is that the economic thinking on this topic has not been well fleshed out in the academic literature. The authors of this paper have set a bold agenda that should provoke considerable further development of our thinking on this important topic.

REFERENCES

Ellison, G., and S. F. Ellison. 2011. ''Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration.'' *American Economic Journal: Microeconomics* 3 (1): 1–36.

Prescott, E., and M. Visscher. 1977. ''Sequential Location among Firms with Perfect Foresight.'' *Bell Journal of Economics* 8 (2): 378–93.

Shapiro, C. 2007. ''Patent Reform: Aligning Reward and Contribution.'' *Innovation Policy and the Economy* 8: 111–56.

Vogel, J. 2008. ''Spatial Competition with Heterogeneous Firms.'' *Journal of Political Economy* 116 (3): 423–66.