Valuing patents and trademarks in complex production chains

Ben Klemens*
Office of Tax Analysis, U.S. Department of the Treasury
ben.klemens@treasury.gov

13 December 2018

Abstract

This article presents a new theoretical framework for evaluating the proportion of a product's surplus attributable to intellectual property (IP, such as patents or trademarks). It is used to explain the empirically observed differences between the use patterns of IP in industries based on discrete products, where patents are typically used to maintain monopolies; versus industries based on complex products, where patents are primarily licensing tools. In a complex production process, learning by doing allows a leading firm to gain some surplus without IP. As the number of steps approaches infinity, the surplus attributable to IP approaches zero. The value of the same patent held by a lagging producer for licensing to the leader does not approach zero, so the ratio of value for licensing to value for maintaining a leading monopoly diverges to infinity. Conversely, for a discrete product, these results do not hold and the traditional use to maintain a monopoly may remain the highest-value use.

Keywords: intellectual property, patent valuation, trademark valuation, production chains, learning by doing
JEL classifications: C65, D450, K110, O340

In a stylized model where an innovation is introduced to an infinite number of producers in perfect competition, competition would drive producer profits to zero both before and after the innovation. By such a model, any and all producer profits could be ascribed to having a right to exclude competitors from use of the innovation.

Correct attribution of profits to intellectual property (IP) is no small concern, given the advent of *IP boxes*, which offer firms a lower rate on profits associated with IP. Griffith et al. [2010] estimate via simulation methods that an IP box lowers IP-associated tax revenue in European countries to between 85% and

*Thanks to John Asker, Sebastien Bradley, Jennifer 8. Lee, Bill Morgan, Mike McDonald, Ryan Nunn, and Neviana Petkova for discussion and commentary.

Electronic copy available at: https://ssrn.com/abstract=2702102

40% of pre-box revenue, depending on the country and assumptions. If an operations company in a high-tax country has a sibling IP holding company in a low-tax country, attributing a percentage of profits to IP allows revenue to be shifted across borders via royalty payments. If firms could use the stylized model to claim that all revenue is IP-related (and examples below will do so), IP-associated tax avoidance using these means is proportionately larger relative to a more limited attribution of revenue to IP.

The primary contribution of this article is a model of IP valuation distinguishing between discrete inventions (drugs based on a single molecule are a typical example) and complex inventions (smartphones or other machines made from hundreds or thousands of parts). The model also allows a distinction between the value of a patent when used by a leading firm to exclude competitors and the value to a lagging firm using the patent for licensing.

Cohen et al. [2000] surveyed 1,478 research and development labs, and found that

> firms commonly patent for different reasons in "discrete" product industries, such as chemicals, versus "complex" product industries, such as telecommunications equipment or semiconductors. In the former, firms appear to use their patents commonly to block the development of substitutes by rivals, and in the latter, firms are much more likely to use patents to force rivals into negotiations.

The stylized model described above is unable to explain these distinctions, leaving little in the theoretical literature to address the differences observed in Cohen et al. and the other empirical articles discussed below; this article presents a model that does accommodate and explain these distinctions. Table 1 summarizes how the value of a patent differs across different uses in different contexts. The bulk of this article (Section 2) focuses on the case of the use of patents to exclude competitors in the context of complex inventions. In this case, as the complexity of the product grows arbitrarily large, the relative value of the patent to a producer approaches zero.

From this key result, two contrasts naturally follow. First, Section 3 considers patents for licensing, and shows that for an arbitrarily complex product, the ratio of licensing value of a patent for a lagging firm to monopoly-maintaining value for a leading firm goes to infinity. Second, for discrete inventions, the number of production steps is not large, so patent value to a leading firm does not approach zero.

A great many patents are for production chain components that are largely valueless when taken out of their production chain context. Even a trademark has no value unless combined with a product or service that can be sold under that trademark. This article first presents some comparative statics describing how the value of the patent as a percentage of overall product value in a single-step, discrete product changes given changes in variables describing learning, then discusses how patents on components of a complex process behave differently from patents on a discrete product. As production becomes more complex, the proportion of value ascribed to the right of a leading firm to exclude others

Electronic copy available at: https://ssrn.com/abstract=2702102

	Prevent imitation	License imitation
Discrete	Value near all surplus	Low value
Complex	Value approaches zero	High value

Table 1: Patents can be used to prevent imitation to maintain a monopoly, or to allow limited imitation via licensing or litigation, and the value of these uses depend on the complexity of the production chain.

from using a component goes to zero. That is, for the elements of a product with an increasing number of elements, there is a "production chain protection" provided by the need for successors of the first mover to implement all components of production, and this production chain protection eventually dominates the value of intellectual property protection on components. In discrete inventions (i.e., those with few or even one component that requires nontrivial learning), the theorems regarding value as the number of steps approach infinity do not apply and patents may retain high value for this purpose.

As discussed in Section 3, Coase's Theorem predicts that the best use of a patent held by a lagging firm is to license its rights to the leading firm. This licensing is often prefaced by litigation, and as the scope of patentable subject matter has expanded into complex fields such as software-based systems, litigation in those fields has expanded, revealing greater value of patents to competitors. The same studies that track this blossoming of litigation in complex fields find no such growth in discrete fields such as pharmaceuticals. Corollary 8 tracks these empirical observations, showing that the ratio of the value of allowance via licensing to the value to a leading firm of using the patent to exclude competitors diverges to infinity as the complexity of production increases.

Outline Section 1 presents a brief overview of some of the primarily empirical literature regarding IP valuation, distinguishing their efficacy in complex versus discrete industries. This section also provides background and motivation for the article: IP valuation is the basis for billions of dollars in tax revenue and market valuation, and there is a good amount of evidence that firms sometimes make wildly inaccurate assessments of the contribution of IP to the value of their products, with a corresponding effect on taxes paid.

Section 2 formalizes the discussion with a model of an intangible whose value increases over time, via learning-by-doing, stronger brand association, or adaptation of other parts of production. The first half develops a measure of value of a patent for a product component relative to the overall value of the product, and presents some basic comparative statics for this value ratio. The second half of this section extends the model from a standalone product to components of a production chain. This extension leads to the key result that the proportion of value from an intellectual property claim in this context approaches zero, which allows substantive distinctions to be made between IP in discrete inventions and IP in complex inventions.

Electronic copy available at: https://ssrn.com/abstract=2702102

This use of patents for licensing to others is discussed in Section 3, on patent thickets. The model in this article will be used to compare the difference in a complex production context between patent value for licensing and value for maintaining a monopoly, providing a theoretical grounding for the empirical observations regarding the prevalence of licensing in industries based on complex products.

1 Background and Literature

In a standard analysis involving perfectly competitive firms, firms have no incentive to innovate. Farrell et al. [2007] summarize the story:

> Generally, in a highly competitive industry without binding capacity constraints, a firm's rewards are relativistic: they stem from being better than its rivals and are not very sensitive to the industry-wide level of unit costs. Thus, if one firm invents a lower-cost production technique that can be adopted by all without paying, no firm benefits much (although consumers do). Thus, neither a participant nor a pure upstream inventor has much incentive to innovate.

That is, firms go from zero profits before a technology is discovered to zero profits after the discovery. This is the economic rationale for patents, because they allow differentiation so that one firm can earn nonzero profits. This will be referred to as *the textbook model* in the discussion to follow, as it does commonly appear in textbooks such as Stiglitz [1993, pp 474–475] and Landsburg [1995, p 359, with profits "much reduced" but not zero].

Observed differences between complex and discrete products

The textbook model implicitly assumes a one-to-one correspondence between patents and products, but the empirical literature shows that there is value in considering products with one or many associated IP claims. *Complex products* are those that require a combination of elements to provide value. At the discrete extreme, a new chemical or drug may be entirely described by one molecule, which can be synthesized using relatively well-known techniques; at the complex extreme, an SEC filing by a defensive patent aggregator reports a count of 250,000 patents that read on to smartphones of 2011 vintage.[1] The aggregator points to all the components that have to function well before a smartphone can be sold: such a high patent count "...can be attributed to the expanded set of features and functionality incorporated in today's smartphones, including touchscreens, internet access, streaming video, media playback, application store readiness and other web-based services, and WiFi connectivity options."

[1] Prospectus of RPX Corp, `http://www.sec.gov/Archives/edgar/data/1509432/000119312511124791/d424b4.htm`

Electronic copy available at: https://ssrn.com/abstract=2702102

Producers have to have some expertise with each of these elements. A quick search of U.S. online retailers will reveal competitors to well-known smartphone and tablet makers like Plum, Ainol, Kocaso, Dragon Touch, Tagital, Nabi, and others. Although available from U.S. retailers such as Amazon.com, they are produced in countries where firms may have less experience with U.S. patent search and licensing. Consumer reviews for these imitating firms are often in line with the problem of learning-by-doing for dozens of components: the product basically works, but has one fatally weak component, like a slow processor, or low-resolution screen, or antennas with bad reception. These products typically sell for about a fifth the price of leading manufacturers' goods.

Complex inventions are not a recent development: Lampe and Moser [2013] counted 4,576 patents for sewing machines in the 1890s. In this context, a trademarked brand is a component of a complex construction, because a brand must be combined with a product or service before it can generate revenue.

The distinction between complex and discrete industries has a long history in the empirical literature, with many authors finding some evidence that patenting is less salient for complex products. Levin et al. [1988] ran a survey through a principal component analysis and found a noticeable difference across industries. They concluded [p 40] that "... policy changes should be assessed at the industry level. For example, in the aircraft industry, ... lengthening the life of patents would tend to have little effect on innovation incentives at the margin. In the drug industry, the effect of a longer lifetime would tend to matter more." From a survey of a hundred firms by Mansfield [1986, p 174], "... the results indicate that patent protection was judged to be essential for the development or introduction of 30 percent or more of the inventions in only two industries—pharmaceuticals and chemicals", yet "... in office equipment, motor vehicles, rubber, and textiles, the firms were unanimous in reporting that patent protection was not essential for the development or introduction of any of their inventions during this period." Bessen and Meurer [2008] estimated the value of patents on "components of complex technologies," based on USPTO classifications. "The mean value [to complex-component industries] is significantly less than the mean value of other patents, although the median value is a bit higher. ... As we might expect, the patents held by chemical firms are much more valuable than those held by other firms." An empirical study by Webster and Jensen [2011] found that "invention owners get some spillover protection from complementary patents embodied in the final product or process." The model of Section 2 provides a theoretical explanation for these empirical results, demonstrating how a complex invention can have "production chain protection" in lieu of or as additional support to legal protection.

Kim and Vonortas [2006, pp 245–246] find that "technological/product complexity of the sector has a positive effect on the propensity to cross license. ... In sectors like electronics, computers and office machines ... the role of patents has been changing in more recent years from an IP protecting instrument to a strategic instrument facilitating deals, exchanges, and alliances." The differences in cross-licensing patterns across industries based on complex versus discrete goods will be discussed in Section 3, along with some additional litera-

Electronic copy available at: https://ssrn.com/abstract=2702102

ture and examples regarding licensing.

Issues with valuation

It is easy to find situations where all profits are attributed to patents. In *Samsung v Apple*, the Court of Appeals for the Federal Circuit (CAFC) found Samsung to be infringing an Apple patent. Samsung's 2016 petition to the Supreme Court for *certiorari* summarizes the damages award:[2]

> ...the Federal Circuit allowed the jury to award Samsung's entire profits from the sale of smartphones found to contain the patented designs—here totaling $399 million. It held that Apple was "entitled to" those entire profits no matter how little the patented design features contributed to the value of Samsung's phones. In other words, even if the patented features contributed 1% of the value of Samsung's phones, Apple gets 100% of Samsung's profits.

The Supreme Court did hear the case, and the ruling held that a patent may apply to a single component of a complex product, or the entire product itself, but made no determination on which is the correct level for consideration before sending the case back to the CAFC (where it it is still being heard as of this writing).[3]

Grubert [2003] found that high R&D companies are more likely to shift revenue from high-tax to low-tax countries. If a gadget maker has one subsidiary selling the gadget in a high-tax country, and a second that owns the IP in a low-tax country, the firm could shift revenue to the low-tax country by overvaluing the royalties owed by the producing subsidiary to the IP-holding subsidiary. The royalty payment might be set to 100% of the profits in the high-tax country, using the textbook model of perfect competition as a rationale.

Several countries offer an IP box (aka patent box) that provides a lower tax rate for income derived from intellectual property, providing another incentive to pass profits to an IP holding subsidiary, converting the revenue from higher-tax operational profits to lower-tax IP revenue [Evers et al., 2014, Klemens, 2017]. For example, in the United States, the 2017 tax reform introduced a preferential rate for overseas income paid to IP held by U.S. companies relative to the rate for overseas income brought in for other operations.

Setting the value of the patent portfolio is only the beginning of the transfer pricing inquiry, but if this is exaggerated (as this article argues a valuation equal to all profits would be) the final conclusion is likely to be as well.

For example, European operations for the online retailer Amazon are centered in a company in Luxembourg, while IP is held by another Luxembourg company (herein Lux SCS) that benefits from a very low IP box rate. A 2014 European Commission investigation regarding the validity of the structure for tax purposes found that Amazon's operations corporation deducts all expenses

[2] `http://www.scotusblog.com/wp-content/uploads/2016/01/15-777_PetitionForAWritOfCertiorari.pdf`, p 2.

[3] `https://www.supremecourt.gov/opinions/16pdf/15-777_7lho.pdf`

Electronic copy available at: https://ssrn.com/abstract=2702102

from revenue, then "the part of the [Amazon operations corp] profit that is not attributed to other functions is paid out to Lux SCS in the form of a royalty" [Almunia, 2014, §66]. Out of $15 billion in 2013 revenue, operations paid $2 billion to the IP holding company, retaining $33 million in taxable profits for itself.

Another European Commission inquiry was made regarding the relationship between furniture maker IKEA's operations corporation in Netherlands (Inter IKEA Systems) and its IP holding company. The inquiry asked whether the license fee was a disallowed transfer of cash to an entity solely for a more favorable tax treatment, rationalized using an argument akin to the textbook model: "The mere legal owner of the [intellectual property] cannot be entitled to receive all the residual profit of the franchise business after paying a limited return to [the operator] for its allegedly routine functions."[4]

Amazon and IKEA are used as examples because details of their arrangements are publicly available in tax investigation reports, but there are abundant comparable stories from an abundance of other firms.

The key result in this article showing that the relative value of patent protection approaches zero in complex industries also allows some policy considerations outside of valuation. The U.S. Congress only has the power to grant patents "to promote the progress of Science and Useful Arts,"[5] so the question of whether patents add value in a given context theoretically has great legal import. Some countries provide multiple tiers of patents, such as the Australian Innovation Patent, which does not require a full examination, provides legal protections with additional caveats, and is in force for eight years instead of the usual twenty.[6] It would be difficult or impossible to draw legal boundaries classifying production chains by complexity, but if multiple tiers of legal protection are available at different costs, applicants with strong production chain protection may self-select into a lower tier of legal protections.

2 The value of a nonexclusive design

This section presents a model to describe the proportion of the total profits from an intangible that are attributable to the right to exclude. Most of the propositions in the first half of this section formalize intuitive results, but they set the stage for the second half, which culminates in Theorem 7: as the complexity of a product with a patent on each step increases, the relative value of the right to exclude competitors goes to zero. But this theorem does not apply to discrete inventions, where the textbook model may apply to indicate that all profits rely on the presence of patent protection. These two statements comprise the first column of Table 1; the second column will be discussed in Section 3.

[4] State aid decision for SA.46470 (2017/NN) — Netherlands, "Possible State aid in favour of Inter IKEA". European Commission, Brussels, 18 December 2017, §165.

[5] U.S. Constitution Art I §8.

[6] http://www.ipaustralia.gov.au/get-the-right-ip/patents/types-of-patents/innovation-patent/

Electronic copy available at: https://ssrn.com/abstract=2702102

First-mover advantage

For an intangible that improves production of goods, there is typically some learning or systemic adjustment required [Arrow, 1962, Gruber, 1994]. Bessen [2015] explains that "a new technology typically requires much more than an invention in order to be designed, built, installed, operated, and maintained. Initially much of this new technical knowledge develops slowly because it is learned through experience...." For example, the first run with a new technique may mostly produce defective products, but over time the producer may find ways to drive the defect rate toward zero, raising productivity accordingly. Case studies by von Hippel and Tyre [1995] described mechanisms for learning by doing, including the discovery of unforseeable problems and information hidden by the complexity of desiging a new system, and recognizing patterns in production.

On the consumer side, products typically go through an adoption phase, often modeled via the Bass diffusion model [Bass, 1969], describing how familiarity with a product and its associated trademark make their way through the population. Modern goods are often networked, meaning that their value increases with the number of other users of the same good [Choi, 1997]; as with learning-by-doing, the value of the final product increases over time, and a new entrant will have to build a new network to expand the value of its new product.

Let the speed of learning (or brand/network development) be given by $s > 0$, where the rate of productivity increase over time is increasing in s. This will be constant for any given production technology, but setting it as a separate parameter will allow discussion of how patent value changes with faster or slower learning.

Let the productivity at time t from a new intangible be a function $p(t, s)$ which is nonnegative for all $t > 0$ and monotonically increasing in t. Let $p(t, s)$ be zero for all $t \leq 0$. Define faster learning to be a higher rate of productivity increase: in the range where $p(t, s) > 0$,

$$\frac{\frac{\partial p(t,s)}{\partial t}}{p(t,s)} = \frac{\partial \ln(p(t,s))}{\partial t}$$

is monotonically increasing in s:

$$\frac{\partial^2 \ln(p(t,s))}{\partial t \partial s} > 0. \tag{1}$$

Past the range where $p(t, s) = 0$, we expect that the learning or adoption curve tapers off, and for any fixed s the rate of productivity increase over time is decreasing in t:

$$\frac{\partial^2 \ln(p(t,s))}{\partial t^2} < 0. \tag{2}$$

To help build intuition, consider the example of $p(t, s) = \left(\frac{t}{t+1}\right)^s$. The top curve of the top plot of Figure 1 shows this curve when $s = 1/2$ (and the curve whose nonzero values begin later in time will be discussed below).

Electronic copy available at: https://ssrn.com/abstract=2702102

If s were zero, learning would be irrelevant and the technology would have productivity one for all $t > 0$; as $s \to \infty$, it takes longer to reach productivity near one. For this example,

$$\frac{\partial \ln p(t, s)}{\partial t} = \frac{s}{t(t+1)},$$

which is increasing in s and decreasing in t, so the conditions in Inequalities 1 and 2 are satisfied.

An analysis of automobile production by Levin [2000] concluded that the learning curve for product quality is better defined directly in terms of time spent and not quantity produced, which provides a motivation for the form above. Also, in a model with time discounting, it is necessary that there be some expression for production at any given time.

Nonetheless, learning by doing is often expressed as a function of cumulative production [Hiller and Shapiro, 1986, Majd and Pindyck, 1989, Mazzola and McCardle, 1996]. Because the production function does not change and there is a monotonic relationship between cumulative production and time, basing the function on time or on quantity will be largely equivalent. The purpose of expressing the function in terms of quantity produced is to make the problem of selecting quantity endogenous. For example, Gabbay [1979] presents the linear programming problem of selecting production in early steps of a chain given that future production steps benefit from more learning-by-doing when earlier steps produce greater quantities. For the results in this article, the choice of production level is irrelevant. Start by assuming that the production level in each period (defined as the period from zero or when patent $i-1$ expires until patent i expires, $i = 1, 2, \dots$) is constant at one widget per period, without regards to the strategic considerations in the above articles. Allowing the endogenous choice of production would lead to different production levels $f_1, f_2, \dots$. Most proofs below will show that some given variable reaches zero in the limit, and multiplying by any finite factor will not change such an outcome, so long as $\epsilon < f_i < \infty, \forall i$ and some fixed ϵ.

As a technical matter, some proofs will require that $d(\int p)/dt = \int(dp/dt)$. Assumption 1 is sufficient for this to be valid; see Casella and Berger [1990, Sec 2.4].

Assumption 1 *The surplus function $p(t, s)$ demonstrates uniform convergence in t and s. A function $p(t, s)$ is uniformly convergent if, for every ϵ, every fixed s, and any sequence $t_1, t_2, \dots \to T$, there exists an n such that $|p(t_k, s) - p(T, s)| < \epsilon$ for all $k \geq n$. Similarly for every ϵ, every fixed t, and sequence $s_1, s_2, \dots \to S$.*

Let r be the time-discounting rate, so a dollar at time t is worth e^{-rt} present dollars. The bottom plot in Figure 1 shows the same sample function used above with time discounting. Goods do not become obsolete in this model, but one could model a technology in a fast-changing market by raising r to accommodate both the lower relative value of a future dollar and the risk of decreased revenue from the obsolescence of the technology. For example, Pakes

Electronic copy available at: https://ssrn.com/abstract=2702102

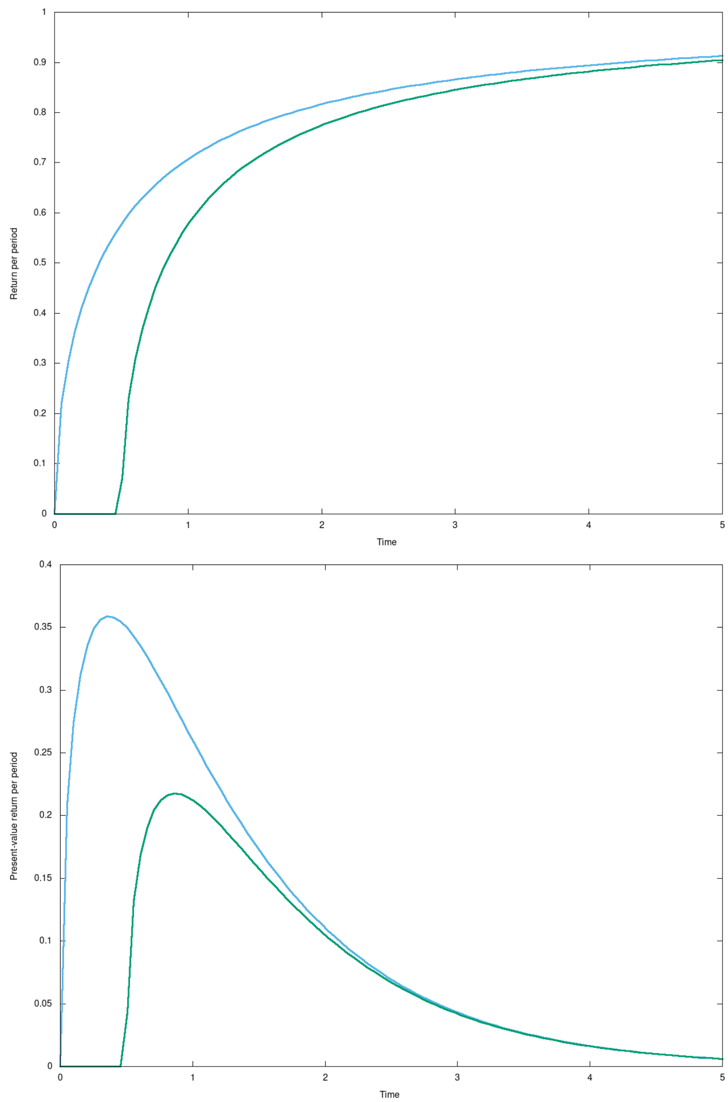

Figure 1: Top: The sample production function for a leading and lagging firm. Bottom: the same production functions time-discounted.

Electronic copy available at: https://ssrn.com/abstract=2702102

and Schankerman [1979] estimated the decay rate in appropriable revenues from a patented technology to be $r = 0.25$, with a confidence interval of $(0.18, 0.36)$.

The total time-discounted value of any continuous convex function $g(\cdot)$, $\int_0^\infty g(t)e^{-rt}dt$, is finite, which guarantees that the total value $p(t, s)$ over all time is finite.

Adding a cost of research to the model will not affect the results in this article. The motivation for explicitly specifying costs in a model is to ask whether total revenue gained by implementing the invention is greater than the cost of researching the invention. This article assumes that the decision has been made to enter, after which the up-front costs are no longer relevant.

Nonetheless, it would be reasonable to ask how firms decide on entry into a market, or other strategic considerations like spending additional fixed costs to speed learning, strategically timing their patents, or trading up-front research spending with learning-by-doing costs. This article will focus on describing a method of evaluating IP and its properties, and leave the use of the method for analysis of strategic options to future research.

After the first entrant adopts the new technology at time $t = 0$, another adopts after some lag $L > 0$. Then the productivity of the late adopter is $p(t - L, s)$, and the time-discounted value of production at time t is

$$V(t, s, L, r) \equiv \begin{cases} p(t - L, s)e^{-rt}, & t > L \\ 0, & t \leq L \end{cases}.$$

With $L = 0$, this equation reduces to the first mover's time-discounted value. The right-shifted curves in the plots in Figure 2 show the value to the leading firm using the sample function, without and with time discounting.

Learning by doing is different from one-time research costs: a competitor with another factory still needs to train workers and adapt its workflow in the same way the first mover did; a competing product needs to work its way up the Bass diffusion curve; a competitor needs to convince customers to switch from the leader's network to their lagging network. As von Hippel and Tyre [1995] conclude from case studies, "it would be very difficult to eliminate doing from learning by doing." One might expect that the lagging firm could gain know-how by hiring employees away from the leading firm, but an analysis of defect and production data from Levitt et al. [2013] found that productivity improvements were less tied to the employees and more to improvements in the production line embodied in adjustments to factory procedures or physical capital, meaning that workers who move to another firm will bring only a fraction of the learning-by-doing gains with them. The empirical literature finds that the spillovers from learning by doing are relatively small. Irwin and Klenow [1994] estimate that "a firm learns over three times as much from an additional unit of its own cumulative output as from an additional unit of another firm's cumulative output". They cite another semiconductor study, Tilton [1971], who states that "much [learning] is either uniquely applicable to a particular operation or can be transferred to another facility only with technical assistance from the firm having the know-how." By these rationales, one would expect that the learning-

Electronic copy available at: https://ssrn.com/abstract=2702102

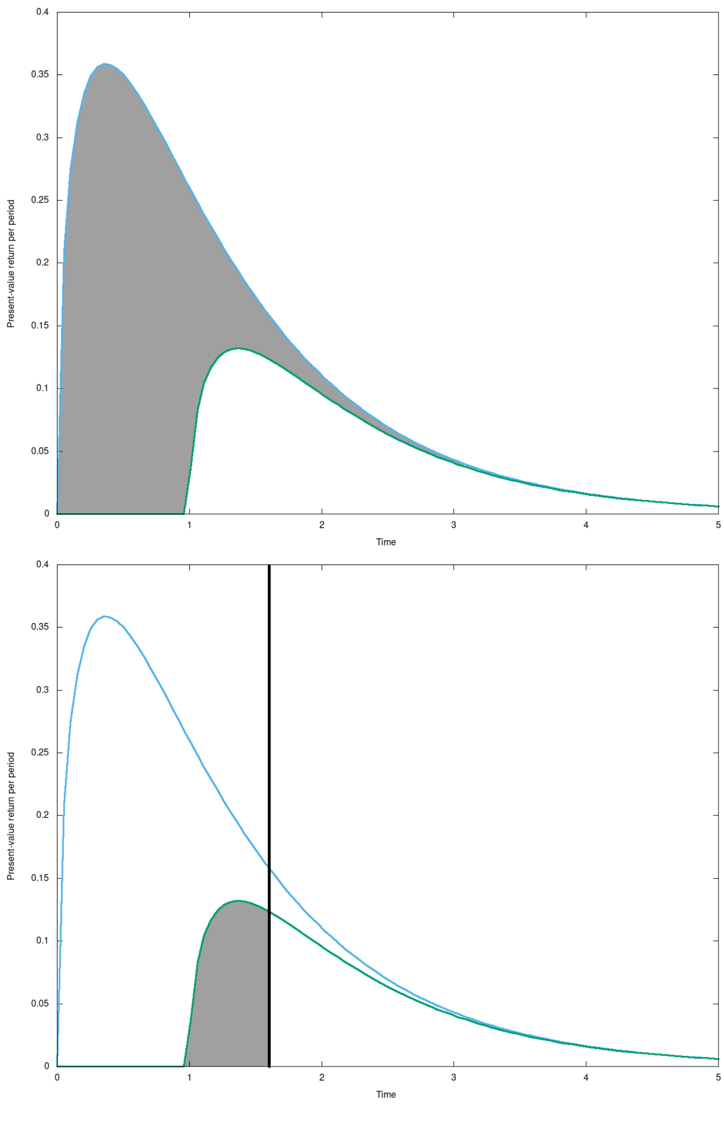

Figure 2: Top: The surplus to the leading firm. Bottom: The value of a patent to the leading firm.

Electronic copy available at: https://ssrn.com/abstract=2702102

by-doing speeds of leaders and laggers would be comparable; the model here simplifies this to the assumption that they are identical.

Regarding surplus gained by the leading firm, this article follows the assumptions of the textbook model of perfect competition: it may cost a monopolist producer using the baseline technology \$100 to produce a widget, but as $p(t, s)$ grows, the producer can make a widget at lower unit costs [Sinclair et al., 2000]. If it has gained enough proficiency to produce widgets at \$90, it has the option to still charge \$100 and make a \$10 profit, motivating the monopolist to invest in technological improvements. But say that other producers have started to improve their processes, and can produce and sell widgets at \$95; then the original producer must now lower costs and sell at \$95 as well, for a \$5 profit. In the limit this leads to the textbook model's conclusion that profits fall to zero after a new technology fully diffuses across competitors. Hiller and Shapiro [1986] assumed that market prices fall over time; these assumptions about a lagged increase in competitor productivity provide a mechanism by which that drop in prices can happen.

Alternatively, learning by doing could leave production costs constant but raise the value of the product to consumers; this is akin to Dosi et al. [2017], who find situations where prices rise with learning leading to quality improvements. Nonetheless, the surplus to the leading firm still relies on the difference in productivity between consumer value added by the leading and lagging firms.

Or, we may assume that productivity expands market share, as for trademarked or other social network-heavy goods. As the competitor improves its product, some portion of the market shifts to the new entrant. The storyline here matches the textbook model as well, as the premium for having the leading network erodes as competitors mature.

These motivating examples assume that the loss of the first mover is equal to the gain of the successor. Modifying the zero-sum assumption of the textbook model to be positive-sum requires additional modeling assumptions and is left for future research. For example, if new competitors add to the network and make the first producer more profitable, then it is actually undesirable to exercise the right to prevent competitors from practicing the invention. If the first producer's production function is an increasing function of the size of the competitors' production, then the result is ambiguous depending on the functional form.

Translating the textbook model's zero-sum assumption to the notation here, the leading firm sees the surplus not taken by the lagging firm:

Assumption 2 *If a lagging firm begins production at time L, the leading firm sees a total surplus of*

$$\int_0^\infty V(t, s, 0, r) - V(t, s, L, r)dt.$$

By definition, there is a natural first-mover lead, such that a competitor can not reproduce the intangible until time L_n, distinct from a right to exclude via legal means (a patent) until time L_p. Assume for now that the patent is absolute

Electronic copy available at: https://ssrn.com/abstract=2702102

and perfectly enforceable, so that no competitors can produce a good until after L_p, although this assumption will be relaxed below. The proofs will assume that L_p is finite and bounded; non-expiring IP protections could be modeled by setting L_p to a large but finite value.

Assume the successors begin the learning process at time L_n, but do not produce a public product until L_p. For example, learning-by-doing with a patented technology could begin before the product can be legally sold. Funk and Magee [2015] catalog many examples of learning by doing before commercialization.

Let the value of the patent be the surplus to the zero-lag firm given a right to exclude minus the surplus to the zero-lag firm given no right to exclude. If $L_p \leq L_n$, then the value of the patent is zero. In the case where $L_p > L_n$, the surplus with a patent beyond the surplus without a patent is the additional profit from not having competitors from time L_n to L_p:

$$\int_{L_n}^{L_p} p(t - L_n, s)e^{-rt}dt. \tag{3}$$

To stress the point, the value of a patent derives from *competitor* productivity blocked by the patent; the productivity of the patent-holding firm does not directly enter into the calculation. The bottom plot in Figure 2 shows the value of the patent for the sample functions.

The value of a patent is decreasing in r, and decreasing in the competitor lag time L_n.[7] For the example function above, Expression 3 is decreasing in s.[8]

However, changes in elements of Expression 3 affect the value of the production process itself as well as the patent. If patent value rises because the surplus value Better would be to consider the ratio of patent value to value without a patent preventing a competitor from extracting value beginning at L_n:

$$R_1 \equiv \frac{\int_{L_n}^{L_p} p(t - L_n, s)e^{-rt}dt}{\int_0^{\infty} p(t, s)e^{-rt}dt - \int_{L_n}^{\infty} p(t - L_n, s)e^{-rt}dt} \tag{4}$$

This is from the perspective of the leading firm, whose profits without patents are still reduced by competition. One could also measure the value of the patent as compared to the total surplus gained by all firms:

$$R_1^A \equiv \frac{\int_{L_n}^{L_p} p(t - L_n, s)e^{-rt}dt}{\int_0^{\infty} p(t, s)e^{-rt}dt} \tag{5}$$

[7] As r rises, e^{-rt} decreases for all values of t. As L_n rises, $p(t - L_n, s)$ decreases, by the assumption that p is increasing in its first term, and the area over which the integral is computed shrinks, reaching zero when $L_n = L_p$.

[8] Proof that $\partial(\int p)/\partial s > 0$: uniform convergence allows us to write this derivative of an integral as the integral of a derivative:

$$\begin{aligned} \partial(\int p)/\partial s &= \int_{L_n}^{L_p} \frac{\partial p}{\partial s} e^{-rt}dt \\ &= \int_{L_n}^{L_p} \ln\left(\frac{t}{t+1}\right)\left(\frac{t}{t+1}\right)^s e^{-rt}dt \end{aligned}$$

which is always negative for $t, s > 0$.

Electronic copy available at: https://ssrn.com/abstract=2702102

One is a function of the other:

$$R_1 = \frac{1}{\frac{1}{R_1^A} - 1},$$

so R_1^A is increasing in some situation iff R_1 is also increasing.

These ratios are well suited to some of the policy questions above, regarding situations where one needs to allocate some percentage of a total surplus (the denominator) to value from patents (the numerator).

Theorem 1 *Assuming Inequalities 1 and 2, and uniform convergence of $p(t, s)$, the ratio of patent value to value without patents (R_1 or R_1^A) is decreasing in s and decreasing in the competitor lag time L_n.*

The proof is largely mechanical, and is relegated to the supplementary materials.

These basic comparative statics formalize the intuition that value can be gained by a first mover even without a legal right to exclude, and that no-patent value increases in absolute terms and relative to patent value as all producers take longer to become proficient with the underlying intangible.

Complex technologies

The remainder of the article considers the case of a product based on two or more production steps.

Assumption 3 *The total productivity is the product of the two subprocesses, each based on a distinct lag time:*

$$V(t, f_1, f_2, L_1, L_2, r) \equiv p_1(t - L_1, f_1)p_2(t - L_2, f_2)e^{-rt}$$

Note that before both lags have passed, one component or the other, and therefore overall productivity, is zero.

For example, the total non-defective rate for a good may be the product of the non-defective rate for the first production step times the non-defective rate of the second step. Roberts [1983] uses a similar rationale as the basis for a model of the rate of learning for a given product. Consumers may first need to decide whether they recognize the brand, then whether they deem the product itself to be valuable, so the total surplus is the product of the likelihood of recognition times the surplus $V(\cdot)$ gained from a sale of the product.

If the surplus added by another step is additive, then we essentially have two separate production processes which could be studied separately. Other monotonic transformations such as logging, exponentiation or multiplication by a constant will preserve most of the results to follow, because the results rely only on production being monotonically increasing in t. Those results that assume a convex production function are irrelevant if a transformation does not preserve convexity.

Electronic copy available at: https://ssrn.com/abstract=2702102

For now, assume that a competing firm begins development of each good as soon as the natural lag passes, and it is unable to produce until the lags on both inputs have passed—in the no-patent case, $\max(L_n^1, L_n^2)$; in the with-patent case, $\max(L_p^1, L_p^2)$. Then Equation 3, the value of a patent to the leading firm, becomes:

$$PV(f_1, f_2, L_n^1, L_n^2, L_p^1, L_p^2, r) \equiv \int_{\max(L_n^1, L_n^2)}^{\max(L_p^1, L_p^2)} p_1(t - L_n^1, f_1) p_2(t - L_n^2, f_2) e^{-rt} dt \tag{6}$$

To reduce notational clutter, this function will be written using a vector of lags $\mathbf{L} \equiv (L_n^1, L_n^2, L_p^1, L_p^2)$

Similar comparative statics to Expression 3 can be verified here. For example, the value of a patent on a compound technology is decreasing in L_n^1 and L_n^2.

Again, changing the production process will affect both the value of the right to exclude and the value of the product with no right to exclude, so consider the ratio of patent value to value without patents:

$$R_2 \equiv \frac{PV(f_1, f_2, \mathbf{L}, r)}{\int_0^\infty V(t, f_1, f_2, \mathbf{0}, r)dt - \int_{\max(L_n^1, L_n^2)}^\infty V(t, f_1, f_2, \mathbf{L}, r)dt}. \tag{7}$$

One could again define a comparison between patent value and total surplus to all producers, R_2^A, by not subtracting the second term in the denominator, and results about changes in R_2 will apply to changes in R_2^A.

If both components in a two-step chain are patented, then the competitor produces nothing between the expiration of the first patent and expiration of the second, and the lead producer gains its full surplus during that period. This is unrealistic. Competitors may be able to "invent around" the patent with an alternative that may have inferior productivity but which still fills the needed step, or may ignore or not be aware of the patent and practice the invention anyway, or may be able to dodge the exclusion by locating in a country with fewer patent enforcement mechanisms, or may use a "knockoff" trademark that some consumers mistake for the leading trademark. A patent has some chance of being held invalid (say, $k\%$), which leads to a probabilistic situation largely equivalent to deterministic partial competition: with $k\%$ odds there will be unrestricted entry by competitors, and with $(1-k)\%$ odds entry is legally restricted. Assumption 4 will accommodate of these real-world considerations, all of which imply some level of productivity by competitors prior to patent expiration.

Assumption 4 *In a model with* partial early production*, for a known sequence of patents, allow competitors to produce C_i percent of their full productivity during the period before expiration of patent i, and reduce $PV(\cdot)$ in this period to $(1 - C_i)PV(\cdot)$.*

Electronic copy available at: https://ssrn.com/abstract=2702102

For sequences of patents discussed below, assume that C_i takes into account the full sequence and is therefore known at time zero.

Assumption 4 accommodates the assumptions made in sequential innovation models such as Scotchmer [1999] or Hopenhayn et al. [2006], where each patent is a discrete step forward in a unidimensional quality space. In those models, a producer who does not have the right to practice the Nth patent but does have rights to the prior $N-1$ patents has a correspondingly lower quality, which in the notation here would be expressed by a nonzero value of C_N.

We would like to compare how one-component and two-component value ratios relate. The comparison will again be the same regardless of the chosen form: $R_2 > R_1$ iff $R_2^A > R_1^A$.

Proposition 2 *Allow partial early production. Assume $L_n^1 < L_p$ (otherwise $R_1 = R_2 = 0$). Then $R_1 > R_2$, and $R_1 - R_2$ is increasing in s_2 and L_n^2.*

The proof is also presented in the supplementary materials.

The surplus to the producer without any competition may rise with the addition of the new component, and the proposition shows that this additional surplus is larger than any additional surplus added by the right to exclude the competitor. The second input to production may not be legally eligible for a patent, such as a law of nature or a web site that does not pass the current tests for patentable subject matter, which one could incorporate into the model by setting $L_p^2 = 0$. Nonetheless, with the relative value of the patent on the first input decreasing with the additional element, relatively more surplus comes from non-patent advantages: surplus including the first intangible gains "production chain protection" by the overall product's dependence on the second component.

If the new component can be imitated immediately and is trivial to learn, the patent and time to learn the first step remains the only bar from competitor entry, but as the second component takes longer to implement, it provides more production chain protection. This provides some intuitive verification for the comparative statics result that the shrinkage in percent of surplus attributable to patents is greater as s_2 or L_2 expand.

A production chain with two steps meets conditions identical to those of a production chain of one step:

Proposition 3 *Let $p^c(t', s') \equiv p(t, s^1, s^2, L^1, L^2)$, where $t' \equiv t - \max(L^1, L^2)$ and s' is implicitly defined by the shape of the amalgamated production function.*

Then in the range where production is nonzero, all the assumptions for a one-unit production function hold for the augmented function: $p^c(t', s')$ is monotonically increasing in t, satisfies Inequalities 1 and 2 for s^1 and s^2, and is uniformly convergent with respect to t, s^1, and s^2.

The proof is given in the supplementary materials.

The ratio of patent value to overall value for the compound production function has comparative statics comparable to that of the discrete production function:

Electronic copy available at: https://ssrn.com/abstract=2702102

Corollary 4 *Let $L_n \equiv \max(L_1, L_2)$. For $p^c(t - L_n, s')$, R_1 is weakly decreasing in s' and L_n.*

The proof: use Proposition 3 to replace $p(t, s)$ in Expression 4 with $p'(t - L_n, s')$, then reapply Theorem 1.

One can chain together repeated applications of Proposition 2. Start with $p_1(\cdot)$ and $p_2(\cdot)$, and use the proposition to show that the ratio based on the first step, R_1^1, is greater than the ratio based on both, $R_2^{1,2}$. Then rewrite the two-step chain as the one-element production process $p_2^c(\cdot)$, which has a ratio of values $R_1^c \equiv R_2^{1,2}$, and use the proposition to show that when combining this with another step $p_3(\cdot)$, we have $R_1^c > R_2^{c,3}$, and so on for each new step in the chain.

There is one complication: the leading firm's compound production function is $p_L^c(t, s') \equiv p_1(t, s) \cdot p_2(t, s)$, and the follower firm's can be expressed after both lags have passed as $p_F^c(t', s') \equiv p_1(t, s) \cdot p_2(t - (L_2 - L_1), s)$. There is no reason to expect p_L^c and p_F^c to have the same functional form, so we need one more step to apply Proposition 2 to R_1 based on a compound function:

Lemma 5 *Rewriting R_1 and R_2 to new ratios R_1' and R_2' which use p_L^c for the leading producer's one-step production function and p_F^c for the following producer's one-step production function, $R_1' > R_2'$.*

With this lemma (proven in the supplementary materials), Proposition 2 can be chained to include an arbitrary number of additional elements, where each new element will reduce the relative value of the patent on the first component.

To this point, we know that adding a production step provides production chain protection, and that protection is increasing as the learning parameter s grows. It is valid to chain these additions, and each will lower the value ascribed to the patent further when s is larger.

The remaining proofs will require a formal definition of a step that requires at least some amount of development.

Assumption 5 *Define*

$$\pi(x) \equiv \int_{\max(L_n^1, L_n^2)}^{L_p} p_1(t - L_n^1, f_1) p_2(x - L_n^2, f_2) e^{-rt} dt$$

and

$$K \equiv \frac{\pi(t)}{\pi(L_p)}. \tag{8}$$

A step has nontrivial learning *for some value ϵ when $K < 1 - \epsilon$.*

The numerator of K is the true value of the patent; the denominator is the value under the counterfactual that production on the second component is constant at its value at time $L_p - L_n^2$ from time $t = 0$ and up. Uniform convergence guarantees nontrivial learning for any given production function, but does not guarantee the condition for an infinite sequence of production functions and any

Electronic copy available at: https://ssrn.com/abstract=2702102

fixed ϵ. The example production function demonstrates this: for any given ϵ, $p(t,s) = \left(\frac{t}{t+1}\right)^s$ will fail the nontrivial learning condition for sufficiently small s. But not all productivity functions converge to trivial learning for small enough s; consider $p(t,s) = \left(\frac{t}{t+1}\right)^{(s+1)}$.

The duration of a patent or other right to exclude is typically time-limited by statute, or even constant, so it is reasonable to assume it is bounded.

Given an infinite number of steps that each involve some learning and some additional conditions, the relative value of the right to exclude is driven to zero:

Theorem 6 *Allow partial early production.*

Consider a production chain with an infinite number of elements, $p_1(t, f_1)$, $p_2(t, f_2), \ldots$. *There is at most a finite number of elements for which the producer is lagging its competitors.*

Assume an infinite number of production components have nontrivial learning for some fixed value of ϵ *shared across all components.*

All patents have a maximum duration L^{max}.

Then the ratio of value from patent protection to no-patent value in Expression 7 goes to zero.

The proof of this result is in the supplementary materials. The proof presents an expression for the shrinkage from adding another step that is bounded below one, so the product of an infinite sequence of such shrinkages approaches zero. As per the comparative statics of Proposition 2, the convergence is faster with larger s or L_n.

This zero relative value result is for one patent, but a product with an infinite number of steps may be achieved using an infinite sequence of patents, which raises the possibility that the sum of patent values converges to a nonzero value. One way this could happen would be if the addition of subsequent patents affects the value of an earlier patent, or there may be details in the infinite sum of infinite sequences. As shown in the proof in the supplementary materials, neither is the case:

Corollary 7 *Assume the conditions of Theorem 6. A producer takes out a sequence of patents at time* $t_p^1, t_p^2, \ldots$ *which expire at times* $t_p^1 + L_p^1, t_p^2 + L_p^2, \ldots$.

Then as the number of (potentially patented) components approaches infinity, the ratio of total value of all patents to overall product value approaches zero.

We thus have the main result: in this theoretical setting with an increasing number of components that require nontrivial learning, the ratio of patent value to non-patent value approaches zero. At the extreme of complexity, only production chain protection is needed for the leading producer to gain the full surplus from production.

Electronic copy available at: https://ssrn.com/abstract=2702102

3 Patent thickets

The discussion to this point has focused on one firm that holds patents, and another that does not make use of any of those patents. But not all patents are practiced by exactly one firm, as some firms license patents and others may (knowingly or not) produce products that infringe a patent. As the number of steps in a production chain approach infinity, the chance that some other party somewhere holds a patent on one of the inputs approaches certainty.[9] This is the origin of the *patent thicket*, wherein no party can produce without cross-licensing at least one step from another party [Shapiro, 2000].

Patent infringement does not require a competitor observing and copying the patent holder. Cotropia and Lemley [2009] give evidence that infringement via independent invention is far more common. Under U.S. law, deliberate copying incurs triple penalties relative to independent invention, so plaintiffs have a strong incentive to allege copying if there is any evidence of copying to be had. But the authors found that "Only 10.9% of the complaints studied... contained even an allegation that the defendant copied the invention, ... copying was established in only 1.76% of all cases in our data set."[10]

There is a simple business strategy based on the frequency of independent invention: obtain a patent, and wait until another party independently invents the covered invention. Baecker [2007] motivates the valuation of patents via option valuation methods [Oriani and Sereno, 2011, Wang, 2011] by pointing out that "patenting has come to resemble the purchase of a lottery ticket."

Galasso and Schankerman [2010] point out that some patents are "weaker" than others, but their definition of weakness is entirely about whether the patent will hold up in court, not whether it is of key significance to an already marketed product. They explain that the best insurance a patent-holder could have for this ambiguity is to bundle patents together and license them as a unit [Choi and Gerlach, 2015]. Such bundles are commonly used to extract royalties from producing firms by entities that do not even produce a product. Depending on the position of the speaker, non-practicing entities following this strategy are referred to as *non-practicing entities* or *patent trolls*. But even actively producing firms may hold a bundle of patents and use a comparable strategy to extract royalties.

Another episode in the patent battle between Samsung and Apple happened in 2013 at the U.S. International Trade Commission, which has the power to block importation of any product manufactured outside the USA. The ITC found that some Apple phones infringed a single patent by Samsung, and declared that those phones were to be blocked from importation to the U.S. en-

[9]This can be formalized: consider a firm that uses k patentable components, each of which has at least a strictly positive chance ϵ of being independently invented. Then the likelihood of some patent being independently invented is $1-(1-\epsilon)^k$, which approaches one as k rises.

[10]In 2011, the America Invents Act created a new defense against infringement (now 35 U.S.C. 273): if a party has been practicing the invention for over a year before the patent was made public, there is no infringement. This law was passed after Cotropia and Lemley [2009], so this exception is not relevant for their study.

Electronic copy available at: https://ssrn.com/abstract=2702102

tirely.[11] The patent in question could have had minimal production chain value, but it gave Samsung potential leverage to extract large rents from Apple before allowing any sale of its products.[12]

In China, a court granted Qualcomm an injunction in December 2018 blocking several types of Apple iPhones from sale, because of two patents: one one photo formatting and one on managing applications using a touch screen.[13] The situation is not yet resolved, but the ruling gives Qualcomm an initial bargaining position similar to Samsung's position in the dispute at the ITC.

Independent invention of a covered invention is not uncommon, and as the number of components that need patent clearance rises, the odds of missing an existing patent rises. Even purchasing a component from a licensed manufacturer is not full protection, as there may be patents in the method of use or combination with other components. If a third party makes a claim against the component maker, under certain conditions U.S. law allows the patent holder to make an infringement claim against the component purchaser (i.e., *vicarious liability*).

Consider a situation where there is a more productive leading firm and a less productive (or not productive) lagging firm, and the lagging firm has obtained a patent on some step in the leading firm's production chain. As per the studies and examples to this point, the most efficient action for the lagging firm is to license the patent to the leading firm and then extract royalties up to the full value of the product. If the leading firm had obtained the patent instead of the lagging firm, it would have produced in exactly the same manner, but without making royalty payments. In a model with differentiated firms, patents thus act as an initial property allocation with no effect on real output, in the style of Coase [1960].

But the rents gained by the leading firm thanks to holding the patent are only the marginal gains from having no competition, whereas the rents that the lagging firm would gain from licensing to the leading firm comprise the full value of the product.

Corollary 8 *Assume a minimum patent length of L_{min}, and the conditions of Theorem 6, including partial early production (perhaps by a third party), and leading and lagging firm as per the definition in Theorem 6. As the number of components in the production chain go to infinity, the ratio of the lagging firm's surplus from holding a patent to the leading firm's surplus from holding the patent goes to infinity.*

Note that the variants of R are a ratio involving no-patent surplus over all time, while this result is about a ratio involving pre-expiration surplus for the

[11] Julianne Pepitone, "Apple banned from selling some iPhones and iPads after Samsung patent win," *CNN tech*. `http://money.cnn.com/2013/06/04/technology/mobile/apple-samsung-itc/index.html`

[12] In the end, the situation was resolved politically: the White House's trade representative interfered and vetoed the ITC ruling.

[13] Associated Press, "Chinese court bans some iPhones over Qualcomm dispute," `https://www.washingtonpost.com/national/chinese-court-bans-some-iphones-over-qualcomm-dispute/2018/12/10/3ebea176-fcae-11e8-a17e-162b712e8fc2_story.html`

Electronic copy available at: https://ssrn.com/abstract=2702102

leading firm. The proof in the supplementary materials therefore needs to take a few steps beyond Theorem 6.

The corollary states that patents are far more useful for licensing in industries based on complex products than those based on discrete products, which reflects empirical findings about how different firms gain from patents. Some of these findings were mentioned in Section 1: the principal component analysis of Levin et al. [1988], the surveys of Mansfield [1986] and Cohen et al. [2000], the USPTO classification study of Bessen and Meurer [2008], the analysis of Webster and Jensen [2011], all of which found salient differences in patent usage by discrete and complex industries.

As per Corollary 8, having a longer production chain simultaneously raises production chain protection for any patented step and raises the chance that a competitor has a patent claim on some other step in the chain. But none of the results regarding a long sequence of production steps have bearing on discrete products. Having a one-step chain, as in a pharmaceutical or many chemical patents, lowers production chain protection while also lowering the risk of competitors blocking other production steps.

Valuing a patent

Because IP has distinct values for prevention of use and allowance of use, an inquiry must be made when valuing a patent as to which use is intended.

For a firm that holds its own IP the valuation may be based on both uses. This may be the scenario when deciding what value to put on a balance sheet for the patent as an asset. Conversely, a license to practice the patent or otherwise use certain associated rights is by definition a partial grant relative to full ownership, and so must have a smaller valuation.

Consider a producing corporation and an IP-holding sibling corporation, where the IP holder is licensing the patent to its sibling. The typical means of pricing for tax purposes is an *arm's length valuation*: value the transfer price under the counterfactual where the IP holder is licensing to an independent company, and use that valuation for the transaction with the sibling. The means of determining that arm's length valuation involves a wealth of legal and economic considerations—and this article adds one more, by arguing that the field of invention has significant weight on the value of the license.

How such a negotiation between an IP holding company an arm's length producer would play out depends on the leverage both parties hold. The holding company of course holds the leverage to shut down the operator entirely. But IP holdings are mere pieces of paper that make no profit by themselves, so a party who can stop production of a product or service based on the IP can eliminate the profits of the IP holder.

For a discrete invention, any two producers have one learning curve to progress along, and may therefore be largely identical. This strips the producers of bargaining leverage, as the IP holding company can leave one producer for another losing only the cost associated with the lag in learning. Under

Electronic copy available at: https://ssrn.com/abstract=2702102

the assumptions of the learning-by-doing curve, as time progresses the cost of switching approaches zero.

Now consider an arm's length negotiation for the IP portfolio covering a complex process, such as the operations of Amazon.com or IKEA. Again, the IP holder can shut down Amazon.com entirely by switching to a competitor, say Nile.com. But if Nile.com is lagging on the implementation of thousands of components, the profitability of the patent is diminished accordingly. In the language of this article, the differentiation that allows nonzero profits are protected not by IP but by the production chain, and this gives Amazon.com operations greater leverage in negotiations.

Valuation for the *sale* of IP is another matter. It is a related question because it is common for the operations company to initially develop the IP, and then sell it to the IP holding sibling. If the sale price matches the time-discounted value of the future royalty stream, such an exchange would be a reasonable shifting of risk from one party to the other, but if the sale price is orders of magnitude smaller than the royalty stream, the exchange seems much more like a tax avoidance scheme. In fact, this article argues that, at the level of abstraction here, the sale valuation should be *higher* than the future revenue stream from production, because IP could have either use to maintain a producing sibling's monopoly or to license to competitors. For a complex portfolio, it is reasonable to expect some components put to each use, and therefore the value of full ownership of the patent to be some combination of the high valuation for licensing and low valuation for own-product protection.

4 Conclusion

Intuitively, for a firm that differentiates itself both by having worked up a learning-by-doing curve and via IP, how profits are ascribed should be split between the production that embodies the learning and the IP. The textbook model is too simple to accommodate this intuitive statement, and the first contribution of this article has been to provide a model where the value of a patent is less than 100% of product value.

This added detail is not merely academic. As per the examples above, the thinking behind the textbook model has been used to justify transfers to IP holding companies that avoid hundreds of millions or even billions of dollars in taxes.

Conversely, the model here demonstrates that any sort of learning or adoption period for an intangible will generate an opportunity for the first mover to see profits above zero, without a legal right to exclude. With a slower ramp-up for both leading and lagging firms, or a longer lag before competitors can adopt the intangible, the value of a patent shrinks relative to the value of the overall product. For discrete products, the value of the right to exclude may still account for the great majority of surplus, but as production expands to arbitrary complexity, production chain protection dominates the value of the right to exclude, which becomes vanishingly small.

Electronic copy available at: https://ssrn.com/abstract=2702102

Such situations of learning-by-doing are common—some authors claim they are universal—and include a ramping-up of productivity via learning-by-doing, a Bass curve of market diffusion, or the development of a network of users. The valuation results are robust to the fact that real-world patents are not an absolute exclusion of competitors, due to substitute products or components, deliberate or unknowing infringement, and the risk that the patent is not enforceable at all.

The article has followed the empirical literature in distinguishing between use of IP for protecting a product and use for licensing. We observe this in real-world complex product based environments like software-based systems [Klemens, 2005, 2008]. Figure 1 summarized the distinctions studies have found across different uses and contexts. The textbook model (and authors that base their assumptions upon the textbook model) states that the valuations for both uses are identical, but this article shows that for complex products they are not only different, but the ratio of the two values can become infinitely large.

For an IP holding company to license to an operating sibling, the valuation for product protection dominates. For the book value of a patent, the value depends on the expected mix of uses between product protection and licensing.

Future directions

In a less theoretical context, how would firms change their tax planning if the textbook valuation allowing the shifting of arbitrarily large profits were not deemed to be valid? One expects that tax havens would be accordingly less attractive, but there may be other changes in behavior as well. Would IP-holding siblings expand licensing to non-siblings? Would the operations sibling simply not transfer some IP to the IP-holding sibling?

The I/O literature on IP traditionally focuses on strategic issues which this article assumed were resolved before the value calculation. But these need not be exogenous, and the valuation method here could be used as a basis for analysis of those considerations. Can a firm spend more up-front research and development costs to speed up learning by doing (i.e., raise s)? Are there reasons to strategically time patents, rather then always get them as soon as feasible? How would a competitor evaluate the entry decision?

Because a single production step could be used to produce a number of products, the model could be used to consider a firm's choices along a tree of production possibilities. The path of innovation was taken as given in this article, but the literature on learning-by-doing typically takes on the problem of which out of a menu of technologies one might choose [Parente, 1994, Jovanovic and Nyarko, 1994, Karp and Lee, 2001, Callander, 2011]. In situations where a right to exclude may be a consideration, the framework presented here may usefully blend with the optimization problems presented in the learning-by-doing literature.

As per the textbook model, the two firms are in a zero-sum game, which may be unrealistic. Firms that hold patents on positive-sum products often license them as parts of standards at a fair, reasonable, and non-discriminatory

Electronic copy available at: https://ssrn.com/abstract=2702102

(FRAND) pricing scheme, where valuations in a learning-by-doing environment may behave differently. The production chains were the same for all producers, but (again depending on design questions) some results could be stated regarding overlapping but distinct production chains.

References

Joaquín Almunia. Alleged aid to amazon by way of a tax ruling. Technical Report SA.38944, European Commission, July 2014. URL `http://ec.europa.eu/competition/state_aid/cases/254685/254685_1614265_70_2.pdf`.

Kenneth J. Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173, 1962.

Philipp N Baecker. An option-based view of imperfect patent protection. *Lecture Notes in Economics and Mathematical Systems*, 587, January 2007. URL `http://ssrn.com/abstract=907983`.

Frank Bass. A new product growth model for consumer durables. *Management Science*, 15(5):215–227, 1969.

James Bessen. *Learning by Doing: The Real Connection between Innovation, Wages, and Wealth*. Yale University Press, 2015.

James Bessen and Michael J Meurer. *Patent failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*. Princeton University Press, 2008.

Steven Callander. Searching and learning by trial and error. *American Economic Review*, 101(6):2277–2308, October 2011. doi: 10.1257/aer.101.6.2277.

George Casella and Roger L Berger. *Statistical Inference*. Duxbury Press, 1990.

Jay Pil Choi. Herd behavior, the "penguin effect," and the suppression of informational diffusion: an analysis of information externalities and payoff interdependency. *RAND Journal of Economics*, 28(3):407–425, Autumn 1997.

Jay Pil Choi and Heiko Gerlach. Patent pools, litigation, and innovation. *The RAND Journal of Economics*, 46(3):499–523, July 2015. doi: 10.1111/1756-2171.12095.

Ronald H Coase. The problem of social cost. *Journal of Law and Economics*, 3:1–44, 1960. URL `http://www.jstor.org/stable/724810`.

Wesley M. Cohen, Richard R. Nelson, and John P. Walsh. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). Working Paper 7552, National Bureau of Economic Research, February 2000. URL `http://www.nber.org/papers/w7552`.

Christopher Anthony Cotropia and Mark A Lemley. Copying in patent law. *North Carolina Law Review*, 87:1421, 2009.

Electronic copy available at: https://ssrn.com/abstract=2702102

Giovanni Dosi, Marco Grazzi, and Nanditha Mathew. The cost-quantity relations and the diverse patterns of "learning by doing": Evidence from india. *Research Policy*, 46(10):1873–1886, dec 2017. doi: 10.1016/j.respol.2017.09.005.

Lisa Evers, Helen Miller, and Christoph Spengel. Intellectual property box regimes: effective tax rates and tax policy considerations. *International Tax and Public Finance*, 22(3):502–530, June 2014. doi: 10.1007/s10797-014-9328-x.

Joseph Farrell, John Hayes, Carl Shapiro, and Theresa Sullivan. Standard setting, patents, and hold-up. *Antitrust Law Journal*, 74(3), 2007.

Jeffrey L. Funk and Christopher L. Magee. Rapid improvements with no commercial production: How do the improvements occur? *Research Policy*, 44(3):777–788, apr 2015. doi: 10.1016/j.respol.2014.11.005.

Henry Gabbay. Multi-stage production planning. *Management Science*, 25(11):1138–1148, 1979.

Alberto Galasso and Mark Schankerman. Patent thickets, courts, and the market for innovation. *The RAND Journal of Economics*, 1(3):472–503, 2010. ISSN 1756-2171. doi: 10.1111/j.1756-2171.2010.00108.x.

Rachel Griffith, Helen Miller, and Martin O'Connell. Corporate taxes and intellectual property: Simulating the effect of patent boxes. Technical Report 112, Institute for Fiscal Studies, 2010. URL `https://www.ifs.org.uk/bns/bn112.pdf`.

H. Gruber. *Learning and Strategic Product Innovation: Theory and Evidence for the Semiconductor Industry*. Studies in Mathematical and Managerial Economics. North Holland, 1994.

Harry Grubert. Intangible income, intercompany transactions, income shifting, and the choice of location. *National Tax Journal*, 56(1, part 2):221–242, March 2003.

Robert Halperin and Bin Srinidhi. U.S. income tax transfer pricing rules for intangibles as approximations of arm's length pricing. *The Accounting Review*, 71(1):61–80, 1996.

Randall S. Hiller and Jeremy F. Shapiro. Optimal capacity expansion planning when there are learning effects. *Management Science*, 32(9):1153–1163, 1986.

Hugo Hopenhayn, Gerard Llobet, and Matthew Mitchell. Rewarding sequential innovators: Prizes, patents, and buyouts. *Journal of Political Economy*, 114(6):1041–1068, December 2006. doi: 10.1086/510562.

Douglas A. Irwin and Peter J. Klenow. Learning-by-doing spillovers in the semiconductor industry. *Journal of Political Economy*, 102(6):1200–1227, 1994.

Electronic copy available at: https://ssrn.com/abstract=2702102

Boyan Jovanovic and Yaw Nyarko. Learning by doing and the choice of technology. Working Paper 4739, National Bureau of Economic Research, May 1994. URL http://www.nber.org/papers/w4739.

Larry Karp and In Ho Lee. Learning-by-doing and the choice of technology: The role of patience. *Journal of Economic Theory*, 100(1):73–92, September 2001. doi: 10.1006/jeth.2000.2727.

YoungJun Kim and Nicholas S Vonortas. Determinants of technology licensing: The case of licensors. *Managerial and Decision Economics*, 27(4):235–249, 2006. URL http://www.jstor.org/stable/25151437.

Ben Klemens. *Math You Can't Use: Patents, Copyright, and Software*. Brookings Institution Press, 2005.

Ben Klemens. The rise of the information processing patent. *Boston University Journal of Science and Technology Law*, 14(1):1–37, Winter 2008.

Ben Klemens. Intellectual property boxes and the paradox of price discrimination. Technical report, Council on Economic Priorities, May 2017. URL www.cepweb.org/intellectual-property-boxes-and-the-paradox-of-price-discrimination/.

Ryan Lampe and Petra Moser. Patent pools and innovation in substitute technologies: Evidence from the 19th-century sewing machine industry. *The RAND Journal of Economics*, 44(4):757–778, 2013. ISSN 1756-2171. doi: 10.1111/1756-2171.12035.

Steven E Landsburg. *Price Theory and Applications*. West Publishing Company, third edition, 1995.

Daniel Z. Levin. Organizational learning and the transfer of knowledge: An investigation of quality improvement. *Organization Science*, 11(6):630–647, 2000.

Richard C Levin, Alvin K Klevorick, Richard R Nelson, and Sidney G Winter. Appropriating the returns from industrial R & D. Working Paper 862, Cowles Foundation for Research in Economics, February 1988. URL http://cowles.yale.edu/sites/default/files/files/pub/d08/d0862.pdf.

Steven D. Levitt, John A. List, and Chad Syverson. Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, 121(4):643–681, 2013.

Saman Majd and Robert S. Pindyck. The learning curve and optimal production under uncertainty. *The RAND Journal of Economics*, 20(3):331–343, 1989.

Edwin Mansfield. Patents and innovation: An empirical study. *Management Science*, 32(2):173–181, 1986. URL http://www.jstor.org/stable/2631551.

Electronic copy available at: https://ssrn.com/abstract=2702102

Joseph B. Mazzola and Kevin F. McCardle. A bayesian approach to managing learning-curve uncertainty. *Management Science*, 42(5):680–692, 1996. doi: 10.2307/2634459.

Alina Ng. *Copyright Law and the Progress of Science and the Useful Arts*. Elgar Law, Technology and Society Series. Edward Elgar, 2011.

Raffaele Oriani and Luigi Sereno. Advanced valuation methods: The real options approach, 2011.

Ariel Pakes and Mark Schankerman. The rate of obsolescence of knowledge, research gestation lags, and the private rate of return to research resources. Working Paper 346, National Bureau of Economic Research, May 1979. URL `http://www.nber.org/papers/w0346`.

Gideon Parchomovsky and R Polk Wagner. Patent portfolios. *University of Pennsylvania Law Review*, 154(1):1–77, 2005.

Stephen L Parente. Technology adoption, learning-by-doing, and economic growth. *Journal of Economic Theory*, 63(2):346–369, August 1994. doi: 10.1006/jeth.1994.1046.

Peter Roberts. A theory of the learning process. *The Journal of the Operational Research Society*, 34(1):71–79, 1983.

Suzanne Scotchmer. On the optimality of the patent renewal system. *RAND Journal of Economics*, 30(2):181—196, Summer 1999.

Carl Shapiro. Navigating the patent thicket: Cross licenses, patent pools, and standard setting. *Innovation Policy and the Economy*, 1:119–150, 2000. URL `http://www.jstor.org/stable/25056143`.

Gavin Sinclair, Steven Klepper, and Wesley Cohen. What's experience got to do with it? Sources of cost reduction in a large specialty chemicals producer. *Management Science*, 46(1):28–45, January 2000. doi: 10.1287/mnsc.46.1.28.15133.

Joseph E Stiglitz. *Economics*. WW Norton & Company, 1993.

John E Tilton. *International Diffusion of Technology: The Case of Semiconductors*. Brookings Institution Press, 1971.

Eric von Hippel and Marcie J. Tyre. How learning by doing is done: problem identification in novel process equipment. *Research Policy*, 24(1):1–12, jan 1995. doi: 10.1016/0048-7333(93)00747-h.

Xiaolu Wang. Patent valuation with a fuzzy binomial model. In *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. Institute of Electrical & Electronics Engineers (IEEE), June 2011. doi: 10.1109/fuzzy.2011.6007345.

Electronic copy available at: https://ssrn.com/abstract=2702102

Elizabeth Webster and Paul H Jensen. Do patents matter for commercialization? *Journal of Law and Economics*, 54(2):431–453, May 2011. URL `http://www.jstor.org/stable/10.1086/658487`.

Appendix

This appendix presents proofs for the main results of the model. It is provided for completeness.

Proof of Theorem 1

This proof demonstrates that the assumptions are sufficient to establish the sign of the derivatives of Expression 4. This has to be done indirectly, because the definition of s is about changes in rates, not the level of $p(t, s)$ itself, which could even be decreasing in s for some range.

Change in s Define:

$$\alpha \equiv \int_{L_n}^{L_p} V(t, s, L_n, r) dt$$
$$\beta \equiv \int_0^{\infty} V(t, s, 0, r) dt$$

With these symbols defined, and using the simpler form excluding competitor profits, the ratio of patent value to no-patent value is

$$R_1^A \equiv \frac{\alpha}{\beta}, \tag{9}$$

or $\ln(R_1^A) = \ln(\alpha) - \ln(\beta)$. We would like to prove that $\partial R_1^A/\partial s$ is negative, which it is iff

$$\frac{\frac{\partial \alpha}{\partial s}}{\alpha} < \frac{\frac{\partial \beta}{\partial s}}{\beta}. \tag{10}$$

It is convenient to restate the integral α to begin at zero:

$$\begin{aligned} \int_{L_n}^{L_p} p(t - L_n, s) e^{-rt} dt &= \int_0^{L_p - L_n} p(t, s) e^{-r(t+L_n)} dt \qquad (11) \\ &= e^{-rL_n} \int_0^{L_p - L_n} p(t, s) e^{-rt} dt \\ &= e^{-rL_n} \int_0^{L_p - L_n} V(t, s, 0, r) dt \end{aligned}$$

Electronic copy available at: https://ssrn.com/abstract=2702102

Because the assumption of uniform convergence guarantees that $d\int V ds = \int dV/ds$, one can similarly restate the derivative of α in terms of the derivative of the value function with respect to s, which will be written as $dV_s(t,s,0,r)$:

$$\partial\alpha/\partial s = e^{-rL_n} \int_0^{L_p-L_n} dV_s(t,s,0,r)dt.$$

Then

$$\frac{\frac{\partial\alpha}{\partial s}}{\alpha} = \frac{\int_0^{L_p-L_n} dV_s(t,s,0,r)dt}{\int_0^{L_p-L_n} V(t,s,0,r)dt}. \tag{12}$$

After the shift, α is the first part of β scaled by e^{-rLn}, so β can be split at $L_p - L_n$ and expressed using α:

$$\begin{aligned} \frac{\frac{\partial\beta}{\partial s}}{\beta} &= \frac{e^{rL_n}\partial\alpha/\partial s + \int_{L_p-L_n}^{\infty} dV_s(t,s,0,r)dt}{e^{rL_n}\alpha + \int_{L_p-L_n}^{\infty} V(t,s,0,r)dt} \\ &\equiv \frac{A+C}{B+D} \end{aligned} \tag{13}$$

Given four expressions A, B, C, and D, where A and B have the same sign, and $A+C$ and $B+D$ have the same sign,

$$\frac{A}{B} < \frac{A+C}{B+D} \quad \text{iff} \quad \frac{D}{B} < \frac{C}{A}. \tag{14}$$

We can use this to show that Expression 12 is less than Expression 13.

For notational convenience define

$$\rho_s(t,s) \equiv \frac{\frac{\partial p(t,s)}{\partial s}}{p(t,s)}.$$

By assumption, $\rho_s(t,s)$ is monotonically increasing in t, so for any $t > L_n - L_p$, $\rho_s(t_1,s) > \rho_s(L_n - L_p, s)$, with the reverse holding for $t < L_n - L_p$. Inserting this constant into numerator and denominator of the D/B ratio shows that it is less than C/A:

$$\begin{aligned} \frac{\int_{L_p-L_n}^{\infty} p(t,s)e^{-rt}dt}{e^{rL_n}\int_0^{L_p-L_n} p(t,s)e^{-rt}dt} &= \frac{\int_{L_p-L_n}^{\infty} \rho_s(L_n-L_p,s)p(t,s)e^{-rt}dt}{e^{rL_n}\int_0^{L_p-L_n} \rho_s(L_n-L_p,s)p(t,s)e^{-rt}dt} \\ &< \frac{\int_{L_p-L_n}^{\infty} \rho_s(t,s)p(t,s)e^{-rt}dt}{e^{rL_n}\int_0^{L_p-L_n} \rho_s(t,s)p(t,s)e^{-rt}dt} \\ &= \frac{\int_{L_p-L_n}^{\infty} dV_s(t,s,0,r)dt}{e^{rL_n}\int_0^{L_p-L_n} dV_s(t,s,0,r)dt} \end{aligned} \tag{15}$$

By Inequality 14, this proves that Expression 13 is less than Expression 12, showing that Inequality 10 is correct, completing the proof that dR/ds is less than zero.

Electronic copy available at: https://ssrn.com/abstract=2702102

Change in L_n Now consider dR/dL_n. The denominator β does not depend on L_n at all, so its derivative does not either. The kernel of α, $V(t, s, L_n, r)$, is decreasing in L_n, and as the lower bound of the integral rises the integral must shrink. Therefore, α and R_1^A unambiguously fall as L_n rises.

Proof of Proposition 2

The proof depends only on $p_2(t)$ being monotonically increasing in t, and does not depend on the second order conditions in Inequalities 1 and 2.

Assume for now no partial early production.

To clarify already complex manipulations, this and subsequent proofs that hold s_1 and s_2 constant will write $p_1(t, s_1)$ as $p_1(t)$, $p_2(t, s_2)$ as $p_2(t)$, and so on. The statement regarding changes in s_2 will be covered in a largely separate discussion after the main proof that the one-step ratio is larger than the two-step ratio.

Let $\Delta \equiv L_n^1 - L_n^2$.

To further simplify the notation, let

$$\begin{aligned}
\zeta &\equiv \int_{L_n^1}^{L_p} p_1(t - L_n^1) p_2(t - L_n^2) e^{-rt} dt \\
&= e^{-rL_n^1} \int_0^{L_p - L_n^1} p_1(t) p_2(t + \Delta) e^{-rt} dt \\
\eta &\equiv (1 - e^{-rL_n^1}) \int_0^{L_p - L_n^1} p_1(t) p_2(t + \Delta) e^{-rt} dt, \\
&= \frac{1 - e^{-rL_n^1}}{e^{-rL_n^1}} \zeta \\
&\text{and} \\
\theta &\equiv (1 - e^{-rL_n^1}) \int_{L_p - L_n^1}^{\infty} p_1(t) p_2(t + \Delta) e^{-rt} dt.
\end{aligned}$$

It is easier to show that $R_2 > R_1$ than $R_2^A > R_1^A$, so we seek to prove that

$$\frac{\int_{L_n^1}^{L_p} p_1(t - L_n^1) e^{-rt} dt}{\int_0^{\infty} p_1(t) e^{-rt} dt - \int_{L_n^1}^{\infty} p_1(t - L_n^1) e^{-rt} dt} \quad \text{(16a)}$$

$$> \frac{\int_{L_n^1}^{L_p} p_1(t - L_n^1) p_2(t - L_n^2) e^{-rt} dt}{\int_0^{\infty} p_1(t) p_2(t) e^{-rt} dt - \int_{L_n^1}^{\infty} p_1(t - L_n^1) p_2(t - L_n^2) e^{-rt} dt} \quad \text{(16b)}$$

The proof will proceed by showing that Expression 16a is greater than $\zeta/(\eta + \theta)$, which is greater than Expression 16b.

Assume for now that $L_n^1 \leq L_n^2$, so Δ is negative and $p_2(t + \Delta) \leq p_2(t)$. Shifting the second term of the denominator in Expression 16b so the integral starts at zero (as in Equation 11) and combining both terms into one integral, then reducing $p_2(t)$ to $p_2(t + \Delta)$, and splitting the integral at $L_p - L_n^1$:

Electronic copy available at: https://ssrn.com/abstract=2702102

$$
\begin{aligned}
& \int_0^\infty \left[p_1(t)p_2(t) - e^{-rL^1_n} p_1(t)p_2(t+\Delta)\right] e^{-rt} dt && \text{(17a)} \\
\geq \quad & (1 - e^{-rL_n}) \int_0^\infty p_1(t)p_2(t+\Delta) e^{-rt} dt && \text{(17b)} \\
= \quad & \eta + \theta && \text{(17c)}
\end{aligned}
$$

Because the numerator of Expression 16b equals ζ, this proves that the expression is less than $\zeta/(\eta+\theta)$.

Now consider the case where $L^1_n > L^2_n$, so Δ is positive.

The sequence of steps from Expressions 18a to 18d below will execute the following steps:

- The procedure until Expression 18b is as above: time-shift the first part of the second integral by L^1_n and use the fact that $p_1(t-\Delta) \leq p_1(t)$ to reduce the first integral and combine it with the second.
- Time-shift the now-smaller expression further back by $-\Delta$.
- With $r\Delta > 0$,

  $$(1 - e^{-rL^2_n})e^{r\Delta} = e^{r\Delta} - e^{-rL^1_n} > 1 - e^{-rL^1_n},$$

  showing that Expression 18c below is greater than Expression 18d.
- For any $g(t)$ and integral of the form $\int_L^\infty p_1(t)g(t)dt$, any value of L less than L^n_1—even negative values—is equivalent, because $p_1(t)$ is zero until $t \geq L^n_1$. This allows us to treat Expression 18d, based on an integral with lower limit $-\Delta$, as an integral with lower limit zero.

These steps arrive at the same inequality relating Expression 16b and Expression 17c.

$$
\begin{aligned}
& \int_0^\infty \left[p_1(t)p_2(t) - e^{-rL^2_n} p_1(t-\Delta)p_2(t)\right] e^{-rt} dt && \text{(18a)} \\
\geq \quad & (1 - e^{-rL^2_n}) \int_0^\infty p_1(t-\Delta)p_2(t) e^{-rt} dt && \text{(18b)} \\
= \quad & (1 - e^{-rL^2_n})e^{r\Delta} \int_{-\Delta}^\infty p_1(t)p_2(t+\Delta) e^{-rt} dt && \text{(18c)} \\
> \quad & (1 - e^{-rL^1_n}) \int_{-\Delta}^\infty p_1(t)p_2(t+\Delta) e^{-rt} dt && \text{(18d)} \\
= \quad & \eta + \theta
\end{aligned}
$$

In both the $L_1 > L_2$ and $L_2 > L_1$ cases, Expression 16b is less than or equal to $\frac{\zeta}{\eta+\theta}$, with equality iff $L^1_n = L^2_n$.

Electronic copy available at: https://ssrn.com/abstract=2702102

The remainder of this proof does not need the assumption that $L_n^1 < L_n^2$ or vice versa, only that $L_p - \Delta > 0$. If this is not the case, then it must be the case that $L_1 > L_p$ and the patent is worth zero.

Define

$$\pi(x) \equiv \int_0^{L_p - L_n^1} p_1(t)p_2(x)e^{-rt}dt.$$

and

$$K \equiv \frac{\pi(t-\Delta)}{\pi(L_p - \Delta)}. \tag{19}$$

For example, given $\int_0^{L_p-L_n^1} p_1(t)e^{-rt}dt$, one could multiply it by the constant $p_2(L_p - \Delta)$ to get $\pi(L_p - \Delta)$. By the monotonicity of $p_2(t)$ in t and the assumption of uniform convergence, $K < 1$.

The sequence of transformations from Expression 20a to 20e below rewrite each subexpression in terms of η, θ, ζ, or K, and thus create a lower bound for Expression 16a. First, time-shift the numerator so the integral begins at zero, time-shift the second integral in the denominator and combine it with the first, split the integral in the denominator at $L_p - L_n^1$, then multiply numerator and denominator by the constant value $p_2(L_p - \Delta)$ so the integrals can be rewritten using $\pi(\cdot)$, and use the monotonicity of $p_2(t)$ to create a lower bound in Expression 20c. By rewriting the expression in terms of K, which we know to be bounded below one, we can add a second inequality which will be used extensively in later proofs.

$$\frac{e^{-rL_n^1}\int_0^{L_p-L_n^1} p_1(t)e^{-rt}dt}{(1-e^{-rL_n})\int_0^\infty p_1(t)e^{-rt}dt} \tag{20a}$$

$$= \frac{e^{-rL_n^1}\pi(L_p-\Delta)}{(1-e^{-rL_n^1})\left(\pi(L_p-\Delta)+\int_{L_p-L_n^1}^\infty p_1(t)p_2(L_p-\Delta)e^{-rt}dt\right)} \tag{20b}$$

$$> \frac{e^{-rL_n^1}\pi(L_p-\Delta)}{(1-e^{-rL_n^1})\pi(L_p-\Delta)+\theta} \tag{20c}$$

$$= \frac{\frac{1}{K}\zeta}{\frac{1}{K}\eta+\theta}$$

$$= \frac{\zeta}{\eta+K\theta}$$

$$> \left[\frac{\eta+K\theta}{\eta+\theta}\right]\frac{\zeta}{\eta+K\theta} \tag{20d}$$

$$= \frac{\zeta}{\eta+\theta}, \tag{20e}$$

which was to be shown.

The definition of partial early production assumes that the full stream of productivity steps are known, so they are unchanging for any step along the

Electronic copy available at: https://ssrn.com/abstract=2702102

chain. That is, both the value of the patent in the one-element production chain and in the two-element chain are reduced by C_1, so the results are retained.

Increasing gap in s_2 and L_s To show the statement that the gap between the one-product ratio and the two-product ratio is larger as s_2 and L_2 grow, first note that the one-product ratio is constant with regards to any characteristic of the second production step. Then, modify the proof of Theorem 1 by replacing $V(t, s, L, r)$ with $V(t, s_1, s_2, L_1, L_2, r)$ in the definitions of α and β, and read L_n to be $\max(L_n^1, L_n^2)$. Because s_1 is constant, the proof using derivatives in s_2 carries through as before. Then the one-product ratio does not change, but the two-product ratio is decreasing in s and L_n^2, widening the gap.

Proof of Proposition 3

Simplify notation by writing $p_1(t - L_1, s_1)$ as $p_1(\cdot)$, and similarly for p^c and p_2.

With both terms of $p^c(t, \dots) = p_1(t, s_1)p_2(t, s_2)$ monotonically increasing in t over the range where production is greater than zero, the whole is monotonically increasing in t.

The log of $p_1(\cdot) \cdot p_2(\cdot)$ is $\ln(p_1(\cdot)) + \ln(p_2(\cdot))$, so

$$\begin{aligned}\frac{\partial^2 \ln(p^c(\cdot))}{\partial t \partial s_1} &= \frac{\partial^2 \ln(p_1(\cdot))}{\partial t \partial s_1} + \frac{\partial^2 \ln(p_2(\cdot))}{\partial t \partial s_1} \\ &= \frac{\partial^2 \ln(p_1(\cdot))}{\partial t \partial s_1}\end{aligned}$$

By assumption, the right-hand side is positive. Similarly for s_2.

Taking derivatives in t,

$$\frac{\partial^2 \ln(p^c(\cdot))}{\partial t^2} = \frac{\partial^2 \ln(p_1(\cdot))}{\partial t^2} + \frac{\partial^2 \ln(p_2(\cdot))}{\partial t^2}.$$

With both terms on the right-hand side assumed to be less than zero, the expression on the left-hand side is also less than zero.

Finally, the product of two uniformly convergent functions is also uniformly convergent.

Proof of Lemma 5

Proposition 3 shows that there is a valid value of s in compound production functions, and having established that there is a valid s, we can again ignore it in cross-production function work and write $p(t, s)$ as $p(t)$.

This proposition can be proven by modifying the proof of Proposition 2. In this modified version, replace $p_2(t)$ with $p_3(t)$ throughout. Do not change ζ, but η and θ will be modified below. Read L_n in the proposition and its proof as L_n^1. Define $\Delta \equiv L_1 - L_3$.

Electronic copy available at: https://ssrn.com/abstract=2702102

What $p_1(t)$ is replaced with depends on whether it is in reference to the leading firm, which now begins with composite function $p_1(t)p_2(t)$, or the lagging firm, whose production is now based on $p_1(t-L_1)p_2(t-L_2)$.

To give an example, here is the original step from Expression 17a to Expression 17b (with $\Delta_{12} \equiv L_1 - L_2$):

$$\int_0^\infty \left[p_1(t)p_2(t) - e^{-rL_n^1}p_1(t)p_2(t+\Delta_{12})\right] e^{-rt}dt$$
$$\geq \qquad (1-e^{-rL_n^1})\int_0^\infty p_1(t)p_2(t+\Delta_{12})e^{-rt}dt$$

Rewrite the first expression as:

$$\int_0^\infty \left[p_1(t)p_2(t)p_3(t) - e^{-rL_n}p_1(t)p_2(t+L_1-L_2)p_3(t+\Delta)\right] e^{-rt}dt \tag{22}$$

and the second as

$$(1-e^{-rL_n})\int_0^\infty p_1(t)p_2(t)p_3(t+\Delta)e^{-rt}dt \tag{23}$$

If $L_1 > L_2$, then the inequality still follows: as before, reduce $p_3(t)$ in the first term of Expression 22 to $p_3(t+\Delta)$ (which was assumed in this case to be negative), but also reduce $p_1(t)p_2(t+L_1-L_2)$ in the second term to $p_1(t)p_2(t)$.

If $L_1 < L_2$, there is some scaling factor $\sigma' < 1$ between the second term of the integral in Expression 22 and the kernel of Expression 23:

$$\sigma'\int_0^\infty p_1(t)p_2(t+L_1-L_2)p_3(t+\Delta)e^{-rt}dt = \int_0^\infty p_1(t)p_2(t)p_3(t+\Delta)e^{-rt}dt$$

Let $\sigma \equiv \max(\sigma', 1)$; then the inequality always holds if Expression 23 is rewritten as

$$(1-\sigma e^{-rL_n^1})\int_0^\infty p_1(t)p_2(t)p_3(t+\Delta)e^{-rt}dt,$$

The same reduction by $(1-\sigma)$ occurs from Expression 18a to 18b, and Expression 20b to 20c.

Therefore, replace η with

$$\eta' \equiv \frac{1-\sigma e^{-rL_n^1}}{1-e^{-rL_n^1}}\eta, \tag{24}$$

and similarly for θ'.

The remainder of the proof proceeds with only the above modifications of notation. For example, the definition of $\pi(x)$ does not depend on any lags at all, so it is unchanged after shifting notation, as is K, which is defined based on $\pi(x)$. Because the $(1-\sigma e^{-rL_n^1})$ subexpressions cancel out, the expression in brackets in Expression 20d is also unchanged:

$$\left[\frac{\eta' + K\theta'}{\eta' + \theta'}\right] = \left[\frac{\eta + K\theta}{\eta + \theta}\right] \tag{25}$$

Electronic copy available at: https://ssrn.com/abstract=2702102

Proof of Theorem 6

As with Proposition 3, larger values of C_s only lower the value of the patent relative to the no-patent value, so if the theorem is proven ignoring C_s, it holds when $C_s > 0$.

We have a sequence of values of R_2, first R_2^2 with production steps $p_1(\cdot)$ and $p_2(\cdot)$, then combining those into a compound $p_c(\cdot)$ and joining this with production step three to produce R_2^3, and so on. We seek to prove that the chained sequence of values $R_2^2, R_2^3, \ldots$ approaches zero as the number of terms in the production process approach infinity.

Inequality 20d shows that, R_1 shrinks to R_2^2 by by a factor of at least the quantity in square brackets,

$$\frac{\eta + K\theta}{\eta + \theta}. \tag{26}$$

The η term is the no-patent return to the competitor before L_p has passed, and θ the return after L_p has passed. The denominator is thus the total no-patent return.

The assumption of nontrivial learning states that $K < 1 - \epsilon$. The remainder of the proof will show that the whole of Expression 26 is bounded by a value strictly less than one.

Equation 25 points out that the primed version of this expression equals the unprimed, so the proof to follow regarding the unprimed version holds for the primed.

Define the competitor's productivity at patent expiration

$$PV_p \equiv p_1(L_p - L_1, f_1)p_2(L_p - L_2, f_2). \tag{27}$$

Each step i may have a distinct but constant and finite value of PV_p.

Then by monotonicity, and the assumption that $L_p \leq L^{max}$, η is bounded above by

$$(1 - e^{-rL_n^1}) \int_0^{L^{max}} PV_p e^{-rt} = (1 - e^{-rL_n^1}) \frac{1 - e^{-rL^{max}}}{r} PV_p,$$

and θ is bounded below by

$$(1 - e^{-rL_n^1}) \int_{L^{max}}^{\infty} PV_p e^{-rt} = (1 - e^{-rL_n^1}) \frac{e^{-rL^{max}}}{r} PV_p,$$

and so, for each step in the process, the ratio is bounded by

$$\frac{\eta}{\theta} < \frac{1 - e^{-rL^{max}}}{e^{-rL^{max}}} \equiv \beta \tag{28}$$

Because $\theta > 0$ for any given step in the chain of added production steps, it is valid to scale Expression 26 by θ, and the result can be bounded by Inequality 28:

$$\frac{\frac{\eta}{\theta} + K}{\frac{\eta}{\theta} + 1} < \frac{\beta + K}{\beta + 1} < \frac{\beta + 1 - \epsilon}{\beta + 1}.$$

This constant is bounded below one.

Electronic copy available at: https://ssrn.com/abstract=2702102

Proof of Theorem 7

How would a subsequent patent affect the productivity of the first patent? By definition, R_1 does not take into account the second production step and its patent at all, and the denominator of both R_1 and R_2, representing no-patent productivity is also unaffected by any part of the patent regime. With non-overlapping patent durations, the numerator of R_2, competitor productivity while the first patent is in force, would not change with the addition of the second patent; with overlapping durations, the numerator would shrink. In sum, the addition of a second patent can only reduce the value of R_2 for the first patent, so if the value of the patent relative to overall product value falls to zero without additional patents, it falls to zero with new patents as well. For clarity, name the limiting value for patent one as component count goes to infinity PR_1.

The same holds for the second patent, so the value of $PR_1 + PR_2 = 0$. The limit of the sequence $\Sigma_i PV_i$ as $i \to \infty$, (that is, the sequence $\{0, 0, 0, \dots\}$) is zero. Again, the addition of partial early production, given that the sequence of $C_s$es is constant and known ahead of time, does not affect the result.

Proof of Corollary 8

The lagging firm can extract royalties equal to the full productivity of the product from the leading firm, from time zero to patent expiration at time L_p. In the notation of the proof of Proposition 2, total pre-expiration value is η (with lags set to zero) and total post-expiration value is θ (with zero lags). Again, the proof carries through identically with η and θ or η' and θ', so the primes will be omitted.

Theorem 6 showed the ratio $\frac{PV_i}{\eta+\theta} \to 0$, but this theorem is about

$$\frac{\eta}{PV_i} = \frac{\frac{\eta}{\eta+\theta}}{\frac{PV_i}{\eta+\theta}}. \tag{29}$$

The proof proceeds by finding a lower bound greater than zero for the ratio of pre-expiration value to total value, $\eta/(\eta + \theta)$. With a bounded numerator and a denominator approaching zero, Expression 29 would go to infinity.

The proof of Theorem 6 defined the value of the production function at patent expiration as PV_p (defined in Expression 27). The production function can be normalized at at each step in the chain to have value one at patent expiration, and the ratios $PV_i/(\eta + \theta)$, η/θ, and $\eta/(\eta + \theta)$ do not change. If the production function at step i is scaled down by the partial early production factor C_i, the normalization still exists, but would be scaled accordingly. The proof depends only on such ratios, so the result holds for the non-normalized values iff it holds for the normalized ratios. All instances of PV_i, η, and θ will be understood to be normalized for the remainder of this proof.

Consider a fictional straight-line productivity function $p(t) = t/L_p$, which the normalizations set to one at $t = L_p$. The values of the straight-line versions

Electronic copy available at: https://ssrn.com/abstract=2702102

of η and θ (herein η_s and θ_s) are found on standard integral tables. Let

$$B \equiv e^{-rL_p}(rL_p + 1);$$

then

$$\eta_s = \int_0^{L_p} \frac{t}{L_p} e^{-rt} dt = \frac{1-B}{r^2 L_p}$$
$$\theta_s = \int_{L_p}^{\infty} \frac{t}{L_p} e^{-rt} dt = \frac{B}{r^2 L_p}$$

By the concavity of productivity over time (Inequality 2), any normalized admissible production function is always greater than the (nonconcave and therefore inadmissible) straight-line productivity function before L_p, and always less than the straight-line function for $t > L_p$. That is, for any admissible η and θ, $\eta_s < \eta$ and $\theta_s > \theta$,

$$\frac{\eta}{\theta} > \frac{\eta_s}{\theta_s} = \frac{1-B}{B},$$

and because L_p is assumed to be bounded below by L_{min},

$$\frac{\eta}{\eta+\theta} = \frac{\frac{\eta}{\theta}}{\frac{\eta}{\theta}+1} > \frac{\frac{\eta_s}{\theta_s}}{\frac{\eta_s}{\theta_s}+1} = 1 - B > 1 - e^{-rL_{min}}(rL_{min} + 1).$$

Because the parameters r and L_{min} are constant and nonzero (and the fact that $e^{-x}(x+1) < 1$ for all $x > 0$) this lower bound is constant and bounded above zero.

The numerator of Expression 29 is now bounded below a positive constant, and the denominator goes to zero, so the ratio goes to infinity.

Because partial early production reduces the value of the patent to the leading firm but does not affect the lagging firm's total licensing value, the result still holds given partial early production.

Electronic copy available at: https://ssrn.com/abstract=2702102