Valuing patents and trademarks in complex production chains

Ben Klemens*

Office of Tax Analysis, U.S. Department of the Treasury

ben.klemens@treasury.gov

May 3, 2018

Abstract

This article presents a new theoretical framework for evaluating intellectual property (IP, such as patents or trademarks), that separates the surplus gained from the IP when used for maintaining a monopoly and when used for extracting rents from competitors. As a product increases in complexity, the ratio of the value of IP for rent extraction to value for maintaining a monopoly diverges to infinity. The model may offer a better guide for policy because it closely reflects the empirically observed differences between the use patterns of IP in industries based on simple versus complex products.

Keywords: intellectual property, patent valuation, trademark valuation, production chains, learning by doing

JEL classifications: C65, D450, K110, O340

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

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.

Conversely, the model presented in this article considers how natural "frictions" of real-world production can allow a leading firm to gain full returns from an innovation without a legal right to exclude.

Although much of the discussion around patents seems to revolve around stand-alone, discrete products, 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 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 from using a component goes to zero. That is, 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 legal protection. 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.

This is in regards to the right to exclude others from use of an invention, giving a leading firm a temporary monopoly on products using the invention.

Conversely, if the right to exclude is held by a lagging firm, it may prefer to license the invention to the leading firm, with fees up to the full value of the product. Corollary 8 shows that the ratio of the value of allowance via licensing to the value of competitor exclusion diverges to infinity as the complexity of production increases. These results clarify the distinction between the two uses of patents, which is also not oft-discussed in the existing patent literature, and why those uses differ in industries based on discrete and complex inventions.

Table 1 summarizes the discussion of this article, mapping out the value of the two types of patent use in discrete and complex industries. The value of the use regarding the allowance of imitation are straightforward and alluded to in the literature, and the value of the patent in an invention with only one component is as per the stylized textbook model. The proof that the use of patents to prevent imitation in complex industries has a value approaching zero, and that the ratio of the lower two cells in the table diverge, are novel contributions of this article.

The results in this article advise reconsideration of policies that give identical treatment to intellectual property (IP) along the full spectrum from discrete to complex goods. There is no 'natural' or 'fundamental' patent value: the same IP may account for almost 100% of the value of one product or almost 0% of the value in another.

To give some examples, at the discrete end of the spectrum typical grocery stores sell generic peanut butter that competes directly against a recognizable brand that sells for a premium, even though in some cases both are even made by the same manufacturer. Without the trademark, the premium good would be unlikely to generate any surplus above the small surplus generic makers see. On the complex end, 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. The leading producers are effectively competing with these lagging firms despite partial or no benefit from many of their patents, because these lesser-known competitors often operate in settings with ineffective patent enforcement. 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, but even despite the low price the leaders still lose a small fraction of the market to the imitators.

Although some of these results may seem obvious, they are already a break from the bulk of the existing literature, where a patent typically has some fundamental value, and a single product realizes that value. To give one example, consider Hopenhayn, Llobet, and Mitchell (2006), who, following Scotchmer (1999), present a model of sequential patents. Each patent is a discrete step forward in a unidimensional quality space, and is a blocking patent for subsequent patents, meaning that subsequent patents can not be implemented in products without the rights embodied in the predecessor patent. Valuing the patent is simply a question of reading the corresponding shift in the quality space, integrated over the time span the patent is leading. There is a direct correspondence between one good and one patent, and the good would see zero profits without the patent. Conversely, the model presented in this article focuses on describing a more nuanced relationship between one product and one patent. Later proofs will allow the case where competitors sell products before patent expiry.

Applications The position that all profits are attributable to patents is sometimes taken in real-world valuations. 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:[1]

> …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).[2]

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 stylized model of perfect competition sketched above as a rationale.

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

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

Several countries offer a "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, Miller, and Spengel 2014; Klemens 2017). For example, in the United States, the 2017 tax reform introduced a *Global Intangible Low-Taxed Income* (GILTI) category giving 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 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).

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."[3]

There are abundant comparable stories, indicating that Amazon and IKEA are not at all unique. They are used as examples only because details of their arrangements are publicly available in tax investigation reports.

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,"[4] 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 good for eight years instead of the usual twenty.[5] 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.

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.

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. After laying the basic groundwork and presenting some comparative statics to help characterize patent value alone and

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

[4] U.S. Constitution Art I §8.

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

as a ratio of overall product value, 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 a patent[6] in this context approaches zero, which allows substantive distinctions to be made between patents in discrete inventions and patents in complex inventions.

This use of patents for licensing to others is discussed in Section 3, on patent thickets. The literature shows a large difference between licensing patterns in industries focused on discrete goods and in industries based on complex goods. 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.

1 Literature

In a standard analysis involving perfectly competitive firms, firms have no incentive to innovate. Farrell, Hayes, Shapiro, and Sullivan (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

[6] The word *patent* originally referred to any government-granted right to exclude (Ng 2011). I streamline language in the remainder of the article by using this definition of *patent*.

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) and Landsburg (1995). This article presents a model in which new technologies are valuable even without a right to exclude.

The model ties in with the literature on complex versus discrete products. *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 that smartphones of 2011 vintage are covered by 250,000 patents.[7] 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." Before the advent of the smartphone, 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 literature, with many authors finding some evidence that patenting is less salient for complex products. Levin, Klevorick, Nelson, and Winter (1988) ran a survey through a principal component analysis and found a noticeable differ-

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

ence 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), "… 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 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 an invention can have "production chain protection" in lieu of or as additional support to legal protection.

Kim and Vonortas (2006) 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.

The question of this article—how much value does a legal right to exclude add to a production chain?—is only one part of the larger question of real-world patent valuation. Much of the patent valuation literature (e.g., Wang (2011)) uses an options model; Oriani and Sereno (2011) describe the binomial options model as "the most commonly applied numerical method." Baecker (2007) motivates the use of option valuation methods by pointing out that "patenting has come to resemble the purchase of a lottery ticket." Others use a market approach or other atheoretical measure of willingness to buy.

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 makes more detailed and perhaps surprising statements about total attributable profits under certain conditions.

First-mover advantage

For an intangible that improves production of goods, there is typically some learning or systemic adjustment required (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.

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.

Such new entrants do exist in the marketplace. For a wide range of complex products, there is a "knockoff" version that is of generally lower quality and lower price, such as the smartphone brands that are almost unknown in the U.S. market. This lower quality version may have a brand with no public recognition (or recognition as inferior), or may be behind on the learning curve relative to the leader. The alternative product may be produced in a location where patents and trademarks are not respected, but sold "under the RADAR" in markets where the leading product is sold, but even while practicing the same patent may still be an inferior product. From the start the knockoffs may strip only a small amount of surplus from the leading producer, but as the lagging firm improves its processes and marketing IP it may pull an increasingly significant amount of surplus away from the leader.

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 $p(t,s) = \left(\frac{t}{t+1}\right)^s$. 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.

As a technical matter, assume that $p(t,s)$ demonstrates uniform convergence in t and s:

Definition 1 *A function $p(t,s)$ is uniformly convergent if, for every ϵ, every fixed s, and any sequence $t_1, t_2, \cdots \to T$, 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 $s_1, s_2, \cdots \to S$.*

Uniform convergence will facilitate some of the proofs by guaranteeing that $d(\int p)/dt = \int(dp/dt)$; see Casella and Berger (1990).

Let r be the time-discounting rate, so a dollar at time t is worth e^{-rt} present dollars. 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 and Schankerman (1979) estimate 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 $p(t, s)$ has a finite integral.

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.

Regarding producer surplus, this article follows the assumptions of the basic textbook model of perfect competition as sketched above: 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. 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 diffuses across competitors.

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 a larger network erodes as competitors mature.

Adding a cost of research to the model will not affect the results in this article. First, the model directly asks how profitable the patented product is, and adding the question of whether that calculated profit is enough to justify an up-front cost adds some complications without new insights. Second, 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.

By these rationales, one would expect that the learning-by-doing speeds would be comparable; the model here assumes them to be identical.

The key consideration for costs in this model is whether the invention will be invented at all, as total revenue needs to be lower than costs. None of the scenarios below will involve a ban on patents, and a claim that patent value relative to patent-free value falls has little bearing on total revenue via one stream or another. As for the competitor, if it has zero up-front research cost, as is assumed in the textbook model, it will always enter.

These motivating examples assume that the loss of the first mover is equal to the gain of the successor. Weakening the zero-sum assumption of the textbook model 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.

Until time L, the sole first mover gains $p(t, s)$ at time t. After the lag of L, the other entrant will enter and gain $p(t - L, s)$ at time t, so following the textbook model's assumptions, the leading firm with zero lag sees 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 L_n periods, distinct from a right to exclude via legal means (a patent) for L_p periods. Assume for now that the patent is absolute 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. Trademarks that eventually lose validity due to genericide are by definition informally used by consumers to refer to competitors well before the trademark becomes unenforceable.

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 value of a patent is decreasing in r, and decreasing in the competitor lag time L_n.[8] For the example function above, Expression 3 is decreasing in s.[9]

However, changes in elements of Expression 3 affect the value of the production process itself as well as the patent. 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:

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

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.

[8] 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, eventually reaching zero when $L_n = L_p$.

[9] 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:

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

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

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, which are currently available at `https://ben.klemens.org/pdfs/klemens-patentval-appendix.pdf`.

This theorem is an intermediate result that formalizes 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. The comparative statics may be valuable by themselves, but the primary intent of this theorem is to set a baseline for comparison to complex production processes.

Complex technologies

Consider the case of a product based on two assets, and assume that the total productivity is the product of the two subprocesses, each based on a distinct lag time: $V(t, f_1, f_2, L_1, L_2, r) = p_1(t - L_1, f_1)p_2(t - L_2, f_2)e^{-rt}$ after both lags have passed, and zero before that point.

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. Consumers may first need to decide whether they recognize the brand, then whether they deem the product itself to be valuable, so the likelihood of purchase is the product of the likelihood of recognition times the likelihood of positive product value of $V(\cdot)$.

If the surplus added by another step is additive, then we essentially have two separate products 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, most of which 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.

Assume for now 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, write this function 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 be 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. A patent is not an ironclad right to exclude: 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 the patent and practice the invention anyway (deliberately or not), or may be able to dodge the exclusion by locating in a country with fewer patent enforcement measures, 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. In all of these cases, the right to exclude is worth less, because it is only partially successful in excluding competition.

Definition 2 *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 $(1-C_i)PV(\cdot)$.*

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

Most of the proofs below show that patent value to the leading firm is small relative to some other value, so the addition of a condition that reduces the value of the patent but has no other effect will only strengthen the results.

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. Given that $L_p^1 > L_p^2$, the ratio of the value of L_p^1 to no-patent value for the discrete technology producing $V(t, f_1, L, r)$ is greater than the ratio for the complex technology $V(t, f_1, f_2, \mathbf{L}, r)$. The change is larger for larger f_2 or larger 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.

Difficulty in learning changes both the value of the patent and the value of the overall product, so it is not immediately obvious that the first shrinks faster than the second as f_2 or L_2 expand. But the math verifies intuition: 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.

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

Corollary 4 *Let $L_n \equiv \max(L_1, L_2)$. For $p^c(t-L_n, s')$, the ratio of patent value to value without patents 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 $p_L^c(t, s') = p_1(t, s) \cdot p_2(t, s)$, and the follower firm's can be expressed after both lags have passed as $p_F^c(t', s') = p_1(t, s) \cdot p_2(t-(L_2-L_1), s)$. There is no reason to expect p_L^c and p_F^c to have the same functional form, so there is some more work to be done (in the proof in the supplementary materials) to extend Proposition 2 to start with a compound function:

Lemma 5 *Rewriting* R_1 *and* R_2 *to new ratios* R'_1 *and* R'_2 *which use* p_c^L *for the leading producer's one-step production function and* p_c^F *for the following producer's one-step production function,* $R'_1 > R'_2$*.*

With this lemma, 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 following proofs will require a formal definition of a step that requires at least some amount of development.

Definition 3 *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 the level it takes at time $L_p - L_n^2$ from time $t = 0$ and up. For a sequence of production steps, the condition of nontrivial learning requires that the ratio of these two integrals does not approach one, which would be the case if the ramp-up to productivity were near-instantaneous. 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 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. Note that 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. It presents an expression for the shrinkage from adding another step, and finds a bound for this expression, which allows the shrinkage at each step to be bounded above zero, so the product of an infinite sequence of such shrinkages approaches zero.

This zero value for one patent may be achieved using an infinite sequence of patents, which raises the possibility that the infinite 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 total value of all patents 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.

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 party somewhere holds a patent on one of the inputs approaches certainty.[10] 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).

Another episode in the patent battle between Samsung and Apple happened in 2013 at the 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 were therefore to

[10] 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.

be blocked from importation to the U.S. entirely.[11] The patent in question could have had zero production chain value, but it gave Samsung potential leverage to extract large rents from Apple before allowing any sale of its products.[12]

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. Cotropia and Lemley (2009) attempt to quantify the rate of independent invention via the claims in infringement cases. 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."[13]

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 serious 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.

[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] 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.

Such bundles are commonly used to extract royalties from competitors. Reback (2002) tells an entertaining story in which IBM negotiated a license from what was at the time a growing spinoff from the Stanford University Network, Sun Microsystems, for a bundle of potentially weak patents:

> My [Sun] colleagues… took to the whiteboard with markers, methodically illustrating, dissecting, and demolishing IBM's claims. We used phrases like: "You must be kidding," and "You ought to be ashamed." Confidently, we proclaimed our conclusion: Only one of the seven IBM patents would be deemed valid by a court, and no rational court would find that Sun's technology infringed even that one.
>
> The chief [IBM lawyer] responded. … "Maybe you don't infringe these seven patents. But we have 10,000 U.S. patents. Do you really want us to go back… and find seven patents you do infringe? Or do you want to make this easy and just pay us $20 million?"

Consider a situation where there is a more productive leading firm and a less 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 leading firm comprise the full value of the patent.

Corollary 8 *Assume a minimum patent length of L_{min}, and the conditions of Theorem 6, including partial early production. The market consists of two firms: the leading firm as per the definition in Theorem 6 and a lagging firm. 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 throughout the paper were a ratio involving no-patent surplus over all time, while this result is about a ratio involving pre-expiration surplus for the leading firm. The proof in the supplementary materials therefore needs to take a few steps beyond Theorem 6.

This result reflects empirical findings about the benefit firms gain from patents.

Some of these findings were mentioned in Section 1: the principal component analysis of Levin, Klevorick, Nelson, and Winter (1988), the survey of Mansfield (1986), 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. To give another example, Cohen, Nelson, and Walsh (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.

As per Corollary 8, having a longer production chain (as in telecom or semiconductors) 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 to zero while also lowering the risk of competitors blocking other production steps to zero.

Valuing a patent Lag time affects patent value: at the extreme, if it takes twenty years for an imitator to gain the network or brand recognition of an existing product, a right to exclude for twenty years has no added benefit. Maintaining the invention as a trade secret may be a better route in this case. Thus, a careful valuation would take into account how long it would take a firm to imitate the patent holder. This is a continuous version of the limiting case where the patent value is zero, as the time taken to reverse-engineer or imitate all elements in a complex good may have a wide range across products.

Because a patent 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 patent, both uses may be intended, and the valuation would be dominated by the larger value of the two. 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.

Now 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 right to allow use via licensing and the right to exclude to maintain a monopoly use are distinct.

To give some background, the initial work to develop the IP is typically done by the operating subsidiary, who then sells it to the IP holding company. If the initial sale of the patents were to equal the value of the expected future revenue stream, the transfer of funds between sibling corporations would be far less of an issue, largely equivalent to a transfer of risk from one party to another. But the initial acquisition price the IP holding subsidiary pays is often drastically low, based on claims that the patent is untested and therefore worth less, changes in legal classification of the transferred intangible, and other such strategies that are beyond the scope of this article. Low initial transfer pricing is what makes it problematic when later payments approach the operating sibling's full profits, as per the examples in this paper's introduction.

In the case of physical property, the value of a license to use a widget-making machine to a producing sibling and to a competitor with a similar production chain will be roughly the same. With only one machine, there is no question of allowing simultaneous use by both sibling and competitor. Even sale of the asset should bear a similar income to that of a perpetual license.

For a patent portfolio, there are many possible arguments for determining the valuation, of which a few will be suggested here. First, the sale to a competitor is a different story from a license of certain rights, so the balance-sheet valuation of the patent is largely irrelevant for valuing a license. As a second option, the IP holder may license to the competitor for the purpose of allowing its operation simultaneous to the producing sibling. Because the IP-holder could conceivably shut down the competitor's operations, the value of the license may approach the full surplus of the competitor.

Third, the IP holder may give an exclusive license to the competitor—that is, the IP holding company allows the competitor to produce and bars the operating sibling from producing. This case, where the threat of shutdown is made toward the operating sibling, puts the patent valuation in the realm of licenses with the intent of excluding imitators, which as per Theorem 6 may have a correct royalty as low as zero for components of a complex product.

Typically, the relationship between a producer and an IP-holding sibling is the third, an exclusive license used to threaten shut-down of the competition. Applying the second scenario to the producing sibling, meaning the IP holder negotiates a price with its sibling based on threat to shut down operation by its own sibling, is an especially far-fetched description of the transaction. If such implicit threats of lawsuits between siblings do exist, they are few and far between. Every situations is different, but this discussion asserts that some high valuations may not be a valid arm's length baseline for IP use whose sole intent is to exclude competitors and the licensor does not have the bargaining position of threatening shutdown of a sibling.

4 Conclusion

A simple classical model of a single good predicts that competing firms see zero profits before an innovation and zero profits after an innovation, unless the first innovator has the right to exclude others from use of the innovation. Under such a model, there is every reason to set the value of the right to exclude equal to any observed profits.

In some situations modeled in this paper, the correct transfer price depends not on the value of the patent in extracting royalties from competitors, but on the value of the patent in preventing competitors from practicing the patent to begin with. The ratio of these two types of valuation diverges as the complexity of the product increases.

This article has discussed some ways in which the value of a right to exclude should typically be lower than the full profits ascribed to an innovation, based on situations where there is a ramping-up of productivity such as learning-by-doing or a Bass curve of market diffusion, or where the intangible under consideration depends on a network of other intangibles such as electronic hardware built from dozens of components or nearly any software (Klemens 2005; Klemens 2008).

Conversely, the model here demonstrates that, even for simple inventions, 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.

The model also shows that in complex technologies, there is a "production chain protection" that allows value to be extracted from an intangible in the chain without reliance on a legal right to exclude. With reasonable assumptions about the addition of new elements to the production chain, a vanishingly small percentage of the value of the overall product is attributable to the right to exclude.

The low valuation in protecting complex production chains from imitators and the high valuation for extracting rents from competitors are not contradictory, but an indication that these are two distinct uses for a patent. This advises that transfer pricing rules should not put all licenses, or discrete and complex products, into the same basket.

Future directions

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.

The intent of the model was to show how much value can be extracted without patents, which in the limit is the full value of the product. This required a number of simplifying assumptions. 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 (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

Almunia, J. (2014, July). Alleged aid to amazon by way of a tax ruling. Technical Report SA.38944, European Commission.

Baecker, P. N. (2007, January). An option-based view of imperfect patent protection. *Lecture Notes in Economics and Mathematical Systems 587*.

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

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

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

Callander, S. (2011, October). Searching and learning by trial and error. *American Economic Review 101*(6), 2277–2308.

Casella, G. and R. L. Berger (1990). *Statistical Inference*. Duxbury Press.

Choi, J. P. (1997, Autumn). 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.

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics 3*, 1–44.

Cohen, W. M., R. R. Nelson, and J. P. Walsh (2000, February). Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). Working Paper 7552, National Bureau of Economic Research.

Cotropia, C. A. and M. A. Lemley (2009). Copying in patent law. *North Carolina Law Review 87*, 1421.

Evers, L., H. Miller, and C. Spengel (2014, June). Intellectual property box regimes: effective tax rates and tax policy considerations. *International Tax and Public Finance 22*(3), 502–530.

Farrell, J., J. Hayes, C. Shapiro, and T. Sullivan (2007). Standard setting, patents, and hold-up. *Antitrust Law Journal 74*(3).

Galasso, A. and M. Schankerman (2010). Patent thickets, courts, and the market for innovation. *The RAND Journal of Economics 1*(3), 472–503.

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

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

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

Hopenhayn, H., G. Llobet, and M. Mitchell (2006, December). Rewarding sequential innovators: Prizes, patents, and buyouts. *Journal of Political Economy 114*(6), 1041–1068.

Jovanovic, B. and Y. Nyarko (1994, May). Learning by doing and the choice of technology. Working Paper 4739, National Bureau of Economic Research.

Karp, L. and I. H. Lee (2001, September). Learning-by-doing and the choice of technology: The role of patience. *Journal of Economic Theory 100*(1), 73–92.

Kim, Y. and N. S. Vonortas (2006). Determinants of technology licensing: The case of licensors. *Managerial and Decision Economics 27*(4), 235–249.

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

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

Klemens, B. (2017, May). Intellectual property boxes and the paradox of price discrimination. Technical report, Council on Economic Priorities.

Lampe, R. and P. Moser (2013). Patent pools and innovation in substitute technologies: Evidence from the 19th-century sewing machine industry. *The RAND Journal of Economics 44*(4), 757–778.

Landsburg, S. E. (1995). *Price Theory and Applications* (Third ed.). West Publishing Company.

Levin, R. C., A. K. Klevorick, R. R. Nelson, and S. G. Winter (1988, February). Appropriating the returns from industrial R & D. Working Paper 862, Cowles Foundation for Research in Economics.

Mansfield, E. (1986). Patents and innovation: An empirical study. *Management Science 32*(2), 173–181.

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

Oriani, R. and L. Sereno (2011). Advanced valuation methods: The real options approach.

Pakes, A. and M. Schankerman (1979, May). 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.

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

Parente, S. L. (1994, August). Technology adoption, learning-by-doing, and economic growth. *Journal of Economic Theory 63*(2), 346–369.

Reback, G. L. (2002, June). Patently absurd. *Forbes*.

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

Shapiro, C. (2000). Navigating the patent thicket: Cross licenses, patent pools, and standard setting. *Innovation Policy and the Economy 1*, 119–150.

Stiglitz, J. E. (1993). *Economics*. WW Norton & Company.

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

Webster, E. and P. H. Jensen (2011, May). Do patents matter for commercialization? *Journal of Law and Economics 54*(2), 431–453.

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:

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

With these symbols defined, and using the simpler form not excluding competitor profits, the ratio of patent value to no-patent value is

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

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:

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

Because the assumption of uniform convergence guarantees that $d\int Vds = \int dV/ds$, one can similarly restate the derivative of α in terms of the derivative of the value function with respect to s, 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 $\exp(-rLn)$, 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 $t > L_n - L_p$, $\rho_s(t_1, s) > \rho_s(L_n - L_p, s)$. 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.

Change in L_n Now consider dR/dL_n. We need to prove that Inequality 10 holds after replacing s with L_n.

Having L_n as a bound to the integral α does not affect the derivative calculations because of the assumption that $p(0,s) = 0$, which implies that $V(L_n, s, L_n, r) = 0$. The steps of time-shifting α and $\partial\alpha/\partial L_n$, breaking β and $\partial\beta/\partial L_n$ into components, and defining $\rho_s(t,s)$ are identical.

By assumption,

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

is monotonically decreasing in t, so

$$\delta_L(t,s) \equiv \frac{\frac{\partial p(t-L_n,s)}{\partial L_n}}{p(t-L_n,s)}$$

is monotonically increasing in t, and Inequality 15 holds.

Proof of Proposition 2

Assume for now no partial early production.

To clarify already complex manipulations, this and subsequent proofs that do no manipulations regarding s 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.

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.

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

To 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

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

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

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 $L_n^1 \le L_n^2$, so Δ is negative and $p_2(t+\Delta) \le p_2(t)$. 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 $L_p - L_n^1$:

$$
\int_0^{\infty} \left[p_1(t)p_2(t) - e^{-rL_n}p_1(t)p_2(t+\Delta)\right] e^{-rt}dt \tag{17a}
$$

$$
\ge (1-e^{-rL_n})\int_0^{\infty} p_1(t)p_2(t+\Delta)e^{-rt}dt \tag{17b}
$$

$$
= \eta + \theta \tag{17c}
$$

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

In the sequence of steps below:

- As per the procedure until Expression 18b is as above: time-shift the first part of the second integral 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.

There are more steps, but they arrive at the same inequality relating Expression 17a and Expression 17c.

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

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 $L^1_n = L^2_n$.

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

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

By the monotonicity of $p_2(t)$ in t, $K < 1$.

Now transform Expression 16a: time-shift the numerator so the integral begins at zero, time-shift the second integral in the denominator and combine it with the first, multiply numerator and denominator by the constant value $p_2(L_p - \Delta)$, split the integral in the denominator at $L_p - L_n^1$, and use the monotonicity of $p_2(t)$ to create a lower bound:

$$
\begin{aligned}
& \frac{e^{-rL_n^1} \int_0^{L_p - L_n^1} p_1(t) e^{-rt} dt}{(1 - e^{-rL_n}) \int_0^{\infty} p_1(t) e^{-rt} dt} && \text{(20a)} \\
= \; & \frac{e^{-rL_n^1} \pi(L_p - \Delta)}{(1 - e^{-rL_n^1}) \left(\pi(L_p - \Delta) + \int_{L_p - L_n^1}^{\infty} p_1(t) p_2(L_p - \Delta) e^{-rt} dt \right)} && \text{(20b)} \\
> \; & \frac{e^{-rL_n^1} \pi(L_p - \Delta)}{(1 - e^{-rL_n^1}) \pi(L_p - \Delta) + \theta} && \text{(20c)} \\
= \; & \frac{\frac{1}{K} \zeta}{\frac{1}{K} \eta + \theta} \\
= \; & \frac{\zeta}{\eta + K\theta} \\
> \; & \left[\frac{\eta + K\theta}{\eta + \theta} \right] \frac{\zeta}{\eta + K\theta} && \text{(20d)} \\
= \; & \frac{\zeta}{\eta + \theta},
\end{aligned}
$$

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 chain. That is, both the value of the patent in the one-element production chain and in the two-element chain are reduced by the same percentage, 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 $\Delta \equiv L_1 - L_3$.

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 one-step 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:

$$\begin{aligned} & \int_0^\infty \left[p_1(t)p_2(t) - e^{-rL_n^1}p_1(t)p_2(t+\Delta)\right] e^{-rt}dt \\ \geq \quad & (1-e^{-rL_n^1})\int_0^\infty p_1(t)p_2(t+\Delta)e^{-rt}dt \end{aligned}$$

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

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

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

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

The same reduction 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:

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

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,

$$\frac{\eta + K\theta}{\eta + \theta}. \tag{25}$$

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. Replacing η and θ with η' and θ', and $(1 - e^{-rt})$ with $(1 - \sigma e^{-rt})$ has no effect, so it is easier to use the pre-prime notation.

The ratio will be shown to have a bound below one, and to be the same bound when η and θ are replaced with η' and θ' to produce the bracketed portion of Expression **??**.

The assumption of nontrivial learning states that $K < 1 - \epsilon$.

Define the competitor's productivity at patent expiration

$$PV_p \equiv p_1(L_p - L_1, f_1)p_2(L_p - L_2, f_2). \tag{26}$$

Each step i may have a distinct (finite) value of PV_p.

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

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

and θ is bounded below by

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

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

$$\frac{\eta}{\theta} < \frac{1 - e^{-rL^{max}}}{e^{-rL^{max}}} \equiv \beta \tag{27}$$

Because $\theta > 0$ for any given step in the chain of added production steps, it is valid to scale Expression 25 by θ, and the result can be bounded by Inequality 27:

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

This constant is bounded below one.

With this constant being the bound for any given step in the production chain, the sequence of constants as $i \to \infty$ converges to the same constant.[14]

This proves the theorem without partial early production: each addition to the production chain reduces Expression 7 by at least this amount, meaning that in the limit it approaches zero.

The value of Expression 7 without partial early production is always greater than the value with, so the expression with partial early production must also approach zero in the limit.

Proof of Theorem 7

How would a subsequent patent affect the productivity of the first patent? Before the addition of a new patent, we would have a ratio of patent value to

[14]This last step regarding the sequence of constants is necessary because we have not said anything about the bounds on L_n^1 or $(s - \exp(-rL_m))$, which may go to (but by the model assumptions can not achieve) zero.

overall product value of R_1, which does not take into account the second production step and its patent at all. The producer's no-competition productivity, $\int_0^\infty p_1(t)p_2(t)\exp(-rt)dt$, is unaffected by any part of the patent regime. If the patent durations do not overlap, only the competitor productivity after the initial patent expiration would be reduced. Adding these two points together, the denominator of R_2, surplus to the producer, expands with the implementation of the second patent. With non-overlapping patent durations, the numerator of R_2, competitor productivity while the first patent is in force, would not change; 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.

As above, the value ratio for the first patent is unaffected by the addition of new patents, and goes to zero. For clarity, let this limiting value be $PR_1 = 0$.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 Theorem 6, 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

$$\frac{\eta}{PV_i} = \frac{\frac{\eta}{\eta+\theta}}{\frac{PV_i}{\eta+\theta}}. \tag{28}$$

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 28 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 26). 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. 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 is also one at $t = L_p$. The values of the straight-line versions 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 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 28 is now bounded above 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.

	Prevent imitation	Allow 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/litigation purposes, and the value of these uses depend on the complexity of the production chain.