Patent Litigation Waves: 1923–2002

Paul Rogerson

Patent Litigation Waves: 1923–2002

Paul Rogerson[*]

The U.S. patent system has experienced several large waves of litigation. However, their origins and implications for innovation remain unclear. A particular challenge is the lack of detailed, long-term data.

This paper builds a new database of patent litigation spanning 1923–2002 by using a large language model to digitize decades of Patent Office records and linking them to a patent-level measure of innovation from Kelly et al. (2021) that scores patents based on novelty (over past patents) and influence (on future patents). A unique feature of this data is that it shows the individual patents asserted in litigation (rather than just the total number of suits).

There are two results. First, technological revolutions have played a large role in explaining waves of patent litigation. Top-scoring, revolutionary patents account for a majority of the variation in litigation volume over time. Second, litigation of revolutionary patents in particular is driven by longevity. These patents continue to be heavily litigated even a decade or more after they leave the Patent Office.

These results help to explain why patent "thickets" emerge in particular decades—the accretion of basic patents during major technological waves creates clusters of long-lived rights that cover the next steps in the technology tree. They also suggest that courts should vary the scrutiny given to litigated patents over time.

[*] Visiting Assistant Professor, Chicago-Kent College of Law. Thanks to Clarivate for access to the LitAlert data.

I. INTRODUCTION

The U.S. patent system has experienced several large waves (or "explosions") of litigation.[1] In fact, we are at the tail end of one right now. Between 1990 and 2000, the per-capita volume of patent lawsuits filed in the U.S. roughly doubled and remained elevated through the 2000s, before gradually beginning to drop in recent years.[2] But historical work shows that patent litigation waves are not a uniquely modern phenomenon. Katznelson (2014) reports that per-capita litigation rates reached similar levels in the 1920s and 1930s (perhaps even a touch higher), before falling in the 1940s.[3] Data is less available for earlier periods, but studies of the busiest individual trial courts of the nineteenth century—including leading work by Beauchamp (2016)—provide very strong reason to believe that the patent system experienced a large wave of litigation in the late nineteenth century, peaking in the 1870s and 1880s.[4]

Each litigation wave has raised concerns that over-abundant patent litigation seemed to be stifling innovation, rather than promoting it. These concerns have often focused on problems of fragmented rights—so-called patent "thickets" or patent "anti-commons."[5] The essential issue is that new technologies are often built on a long chain of discoveries, and therefore are covered by multiple patents. Think, for example, of a modern cell phone. There is the screen, the chips, the software, the antenna, and many others. In total, the final product may be covered by hundreds or thousands of separate patents. This is not just an issue in modern technology. When radio was first developed in the early twentieth century, the inventor of FM radio reported that "it was absolutely impossible to manufacture any kind of workable apparatus without using practically all of the inventions which

---

[1] James Bessen & Michael J. Meurer, *The Patent Litigation Explosion*, 45 LOY. U. CHI. L.J. 401 (2013); Ron D. Katznelson, *A Century of Patent Litigation in Perspective* (2014), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2503140; Christopher Beauchamp, *The First Patent Litigation Explosion*, 125 YALE L. J. 848 (2016).

[2] *See* Bessen & Meurer, *supra* note 1 (rise in the 1990s). A change in the joinder rules for patent cases in 2011 (limiting the joinder of multiple defendants) produced a mechanical increase in the volume of suits. But an appropriate normalization—such as looking at the number of unique patents asserted—shows that the volume of suits began to fall in the late 2010s.

[3] Katznelson, *supra* note 1.

[4] Beauchamp, *supra* note 1.

[5] *See, e.g.,* Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698 (1998); Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, *in* 1 INNOVATION POLICY AND THE ECONOMY 119, 119 (Adam B. Jaffe, Josh Lerner & Scott Stern eds., 2001).

were then known."[6] And locomotives—a key technology in the late nineteenth century—were covered by a "complex web" of overlapping patents on the wheels, the brakes, the engine, the cars, and other components.[7]

Fragmented patent rights create uncertainty and legal friction. Any company that wants to make the final product must locate all relevant patent-holders, strike bargains to license the patents, and worry about surprise lawsuits from hold-outs. Indeed, it may be rational for patent-holders to strategically lay in wait until companies make fixed investments in product design, and then spring with a lawsuit.[8]

Each major wave of patent litigation wave has been a historical period where this problem seemed to bite particularly hard. During the first major litigation wave in the nineteenth century, the Supreme Court argued that patent litigation had "embarrass[ed] the honest pursuit of business with fears and apprehensions of concealed liens and unknown liabilities to lawsuits."[9] In the early twentieth century, legal commenters complained that "conflicting, legally enforceable rights [made] for confusion and obstruction" in the development of new technologies.[10] And modern observers of the most recent patent litigation wave have described a "dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology."[11]

But what is going on remains something of a mystery—why does the problem of fragmented rights seem to flare up and then fade away every few decades?

---

[6] Gaston Llanes & Stefano Trento, *Anticommons and Optimal Patent Policy in a Model of Sequential Innovation*, Harvard Business School Working Paper No. 09-148 (2009), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1424498.

[7] See Steven W. Usselman, *Patents Purloined: Railroads, Inventors, and the Diffusion of Innovation in 19th-Century America*, 32 Tech. & Culture 1047, 1063 (1991); Steven W. Usselman, REGULATING RAILROAD INNOVATION: BUSINESS, TECHNOLOGY, AND POLITICS IN AMERICA, 1840–1920, at 104–10 (2002); Christopher Beauchamp, *The First Patent Litigation Explosion*, 125 Yale L.J. 848, 929–30 (2016).

[8] *See, e.g.,* Erik Hovemkamp, Jorge Lemus, & John L. Turner, *Product Development with Lurking Patentees*, 56 RAND J. ECON. 163 (2025); Carl Shapiro & Mark A. Lemley, *The Role of Antitrust in Preventing Product Holdup*, 168 PENN. L. REV. 2019 (2020).

[9] *Atlantic Works v. Brady*, 107 U.S. 192, 200 (1883).

[10] Sylvester Petro, Patents: Judicial Developments and Legislative Proposals, 12 U. CHI. L. REV. 352, 390 (1945).

[11] Shapiro, *supra* note 5 at 120.

A natural guess is that this has something to do with technological progress.[12] Perhaps major waves of patent litigation follow behind major waves of technological progress, and these bursts of technological advance somehow explain the problem of divided rights. But, in practice, this idea has worked much less well than it seems like it should. So far, the most detailed studies of patent litigation—using granular litigation data—have not found clear evidence that waves of patent litigation are explained by technological revolutions.[13]

However, an important challenge has been the lack of data. Truly granular litigation data—that shows details like the individual patents asserted in each suit—is only available for a handful of years, many of them recent. And a related challenge is the lack of a long-term, quantitative measure that can link individual litigated patents to broader technological waves. These limitations have made it difficult to trace the origins of patent litigation waves, and to see what is going on inside them.

This paper has a new approach with two parts. First, it builds a new database of patent litigation by using a large language model to digitize decades of Patent Office records. These records come from a 1922 amendment to the patent statute that required clerks of court to send litigation notices to the Patent Office each time a case was filed.[14] The notices were published in the Patent Office's Official Gazette from 1923–1984. Post-1984 (and for a few overlapping years in the early 1980s), the information from the notices is contained in the commercial database LitAlert.[15] A unique feature of this data is that it shows the individual patents asserted in each case—the notices are required by law to contain this information.

Second, the paper links the litigation database to a new quantitative measure of the novelty and influence of every U.S. patent issued between 1840 and 2002 from Kelly et al (2021).[16] This measure scores issued patents based on two characteristics: (1) whether they are textually *dissimilar* from earlier patents (novel), and (2) whether they are textually *similar* to future patents (influential). For example, the first patent to use the phrase "alternating current" would score highly by this metric. That phrase is novel (does not appear in prior patents), but also

---

[12] *See* Katznelson, *supra* note 1; B. Zorina Khan, *Innovations in Law and Technology, 1790–1920*, in 2 *Cambridge History of Law in America* 483 (2008).
[13] *See* Bessen & Meurer (2013), *supra* note 1; Beauchamp (2016), *supra* note 1.
[14] R.S. § 4921 (1922).
[15] Thanks to Clarivate for access to the LitAlert data.
[16] Bryan Kelly et al., *Measuring Technological Innovation over the Long Run*, 3 AM. ECON. REV.: INSIGHTS 303 (2021).

influential (does appear in many future patents that build on and apply the breakthrough). Intuitively, this metric can capture basic new ideas entering the patent system. The authors break patents into deciles (top 10%, next 10%, and so on), and construct a long-term innovation index by using the number of top-decile patents issued by the Patent Office in each year.

In total, the linked litigation database spans 1923–2002. The paper reports two core results.

First, technological revolutions have played a large role in explaining waves of patent litigation. The Kelly et al. (2021) innovation index reports three large waves of basic innovation during the period 1840–2002: one in the late nineteenth century (featuring patents on manufacturing, transportation, and electricity), another in the 1920s and 1930s (featuring patents on chemicals, electronics, and transportation), and a final wave in the late twentieth and early twenty-first centuries (featuring patents on computers and biotechnology).[17] Each of these major waves of basic innovation coincides with one of the large waves of patent litigation that has been identified in U.S. history.

Moreover, waves of basic innovation and patent litigation do not merely coincide—the litigation database shows that (at least during 1923–2002) top-scoring, revolutionary patents are highly litigated and can explain a large share of litigation surges. Quantitatively, top-decile patents by themselves explain more than 30% of the variance in the volume of litigated patents over time, and the top three deciles explain more than 60%. These results imply that the patent system has tended to experience a relatively steady flow of litigation over incremental advances (lower-scoring patents), punctuated by large, periodic surges of litigation over revolutionary technologies (top-scoring patents).

Second, the data also provide insight into a key feature of top-scoring patents that drives litigation—their longevity. A surprising finding is that top-scoring, revolutionary patents are no more likely than any other patent to be litigated in the first year after they issue from the Patent Office. Instead, the difference in litigation rates is driven by longevity—a top-decile patent is much more likely (2-4x) to be litigated 5, 10, or even 15 years after it issues. On reflection, this finding makes sense. Top-scoring patents score highly precisely because they are influential—they contain ideas that become part of the basic technology tree for future innovation. Accordingly, they are more likely to still be used by future products even 5, 10, or 15 years later.

---

[17] *Id.*

These findings suggest a story about why patent "thickets" form in particular decades, and why they are connected to technological revolutions. The key is the behavior of top-scoring patents. These influential ideas cover future generations of technology. During a major wave, many of them issue around the same time. And, as they accrete, form a thicket of long-lived rights that covers the next steps in the technology tree. A cell phone in 2005 might not merely infringe patents issued in the last few years, but a very wide (and possibly hard-to-identify) set of patents issued during the last fifteen. The same could be said of a radio set in 1935, or a train engine in 1885.

These points make a basic case for courts to vary their approach to patent law over time. Courts are charged with supervising patent litigation, and must pass judgment on the validity (and scope) of litigated patents. Any time a court invalidates a patent, there is some risk of error. But the costs on the other side of the ledger can look quite different. During periods of slow technological progress, courts can more freely err on the side of upholding even somewhat doubtful patents, trusting that market forces and licensing negotiations can sort things out, and that, while letting a few bad patents through might over-reward their holders, it is unlikely to have broader impacts on innovation. But that logic holds less well during a technological revolution, when the patent system is running hot. The crowding of patents around basic new ideas tends to create frictions that undermine market solutions, so there is a greater role for courts to wade in and clear the thickets by making assessments about which patents were really worth granting

The rest of the paper is organized as follows. Part II explains basic features of U.S. patent litigation. Part III describes the litigation database that the paper constructs. Part IV carries out the analysis. Part V discusses the results and implications. A brief conclusion follows.

II. PATENT LITIGATION

This Part briefly describes the basic structure and features of U.S. patent litigation (Part A), and what is known about its volume over time (Part B).

A. Basics

The power to create the patent system comes from the Constitution, which grants this power to Congress to encourage

progress in the "useful Arts."[18] The first patent statute was passed in 1790[19] and has been revised from time to time since then, but the core features of the patent system have been relatively stable.

Under the patent statute (as interpreted through the gloss of judicial doctrine), an inventor can obtain a patent for an invention that meets a set of basic conditions. First, the invention must be the *kind* of idea eligible for patenting.[20] Patents will not be granted on laws of nature, natural products, or abstract ideas, but rather are reserved for practical applications.[21] Second, the idea must be literally new—an idea already known to the public cannot be patented.[22] Third, on top of literal novelty, the idea must also clear a threshold of creativity. Modern doctrine calls this the requirement of "nonobviousness," and provides that a patent will not be granted for an idea that would have been obvious to an ordinary worker in the field of the invention at the time the invention was made.[23] Fourth, the inventor must fully and accurately disclose the idea in the patent document (and claim a property right no greater than what has been disclosed).[24]

An inventor obtains a patent by filing an application with the Patent Office, which reviews the application for compliance with the statutory conditions. Review by the Patent Office is a relatively coarse filter.[25] The Office receives a large volume of applications relative to its staff of examiners, so examiners are only able to spend a few hours reviewing each one. Given limited resources, the Patent Office tends to grant patents in close (or even debatable) cases, relying on courts to take a closer look if the patent is ever enforced. Arguably this is the right approach—the vast majority of patents cover inventions that prove to be worthless, so it is efficient to wait to see which patents are enforced, and thoroughly review just those ones.[26] But it does mean that courts tend to be the final decision-makers in cases where there is room for debate about the breadth or validity of a patent.[27]

---

[18] U.S. Const. Art. I Sec. 8.
[19] *See* P. J. Federico, *Operation of the Patent Act of 1790*, 85 J. PAT. & TRADEMARK OFF. SOC'Y 33 (2003).
[20] 35 U.S.C. § 101.
[21] *See Alice Corp. v. CLS Bank Int'l*, 573 U.S. 208 (2014).
[22] 35 U.S.C. § 102
[23] 35 U.S.C. § 103. For a history of older standards, see Edmund W. Kitch, Graham v. John Deere Co.*: New Standards for Patents*, 1966 SUP. CT. REV. 293.
[24] 35 U.S.C. § 112
[25] Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495 (2001); Michael D. Frakes & Melissa F. Wasserman, *Irrational Ignorance at the Patent Office*, 72 VAND. L. REV. 975 (2019).
[26] Lemely, *supra* note 25; *but see* Frakes & Wasserman, *supra* note 25.
[27] Recent statutory reforms create another channel for post-issuance review by

Once issued, a patent gives the holder the right to exclude anyone else from practicing the invention during the term of the patent.[28] Importantly, a patent does not confer any *affirmative* right to practice the invention. Rather, it confers a *negative* right to stop others from doing so. Accordingly, in the case of a complex product (such as a cell phone) that is covered by multiple patents, separate permission from each patent holder must be obtained to make the product.

For most of U.S. history, the length of the patent term was set at 17 years after the date the patent issues from the Patent Office.[29] In the 1990s, term reform switched the term to 20 years after the date the application was filed in the Patent Office.[30] But these often come to nearly same thing, since applications generally take a few years to wind their way through the Patent Office, and the applicant does not obtain enforceable rights until the patent actually issues.

Lastly, if another party uses the patented invention during the term—"infringes" the patent—the remedy is a lawsuit in federal district court.[31] In this lawsuit, the plaintiff has the burden to prove infringement, and the defendant has the right to raise the defense that the asserted patent is invalid, i.e., fails a statutory condition.[32] If the plaintiff prevails in the lawsuit by establishing infringement and defeating any challenges to the patent's validity, the available remedies are damages or an injunction.[33]

B. Volume Over Time

For the twentieth and twenty-first centuries, the most comprehensive source of data about the volume of patent litigation is the Administrative Office of U.S. Courts. This series reports the total number of patent suits filed each year (but not details like the individual patents asserted in each case). This data is collected based on the classifications that civil cases are given when filed, and is likely near-complete.[34] Reporting of patent cases begins in

---

allowing a patent defendant (or any other third party) to petition the Patent Office to take a second look at the validity of an issued patent. *See* 35 U.S.C. §§ 311–319 (*inter partes* review); *see also* Pub. L. 96-517 (*ex parte* reexamination); Pub. L. 106-113 (*inter partes* reexamination).

[28] 35 U.S.C. § 271.

[29] *See* Simon Lester & Huan Zhu, *Rethinking the Length of Patent Terms*, 34 AM. UNIV. INT'L L. REV. 787, 791–793 (2019).

[30] *Id.*

[31] 35 U.S.C. § 281.

[32] 35 U.S.C. § 282.

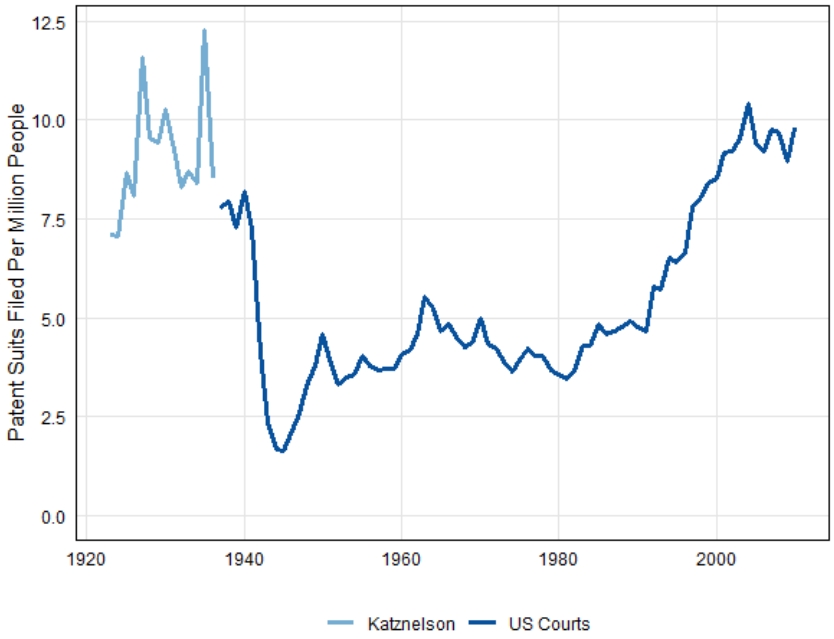
[33] 35 U.S.C. §§ 283–84.

[34] Katznelson, *supra* note 1

1937, and the data is organized by fiscal year, e.g., the data for 1937 covers the period July 1, 1936 to June 30, 1937.

Work by Katznelson (2014) extends the series further back by looking to the litigation notices published in the Gazettes and hand-counting the total number of suits filed each year from 1923 to 1936.[35] As Katznelson notes, there is likely some under-reporting in the Gazettes; Katznelson estimates the reporting rate at perhaps ~87%.[36] The extended series from 1923–2010,[37] scaled per capita and adjusted for under-reporting, is shown in Figure 1 below.

Figure 1: Patent Litigation 1923–2010



The major takeaway is that there have been two periods of elevated patent litigation during the last century—one recently, in the 1990s and 2000s, and another earlier in the twentieth century, in the 1920s and 1930s. These data also match contemporary

[35] Katznelson does not report the individual patents associated with each suit. However, even counting just the total number of suits by hand is a daunting task, not only due to the volume but also because the Gazettes contain a non-trivial number of duplicate entries spread across multiple months (or even years). Even so, Katznelson's suit counts are in remarkably close agreement with this paper's estimates, differing by less than 5% for almost all years.

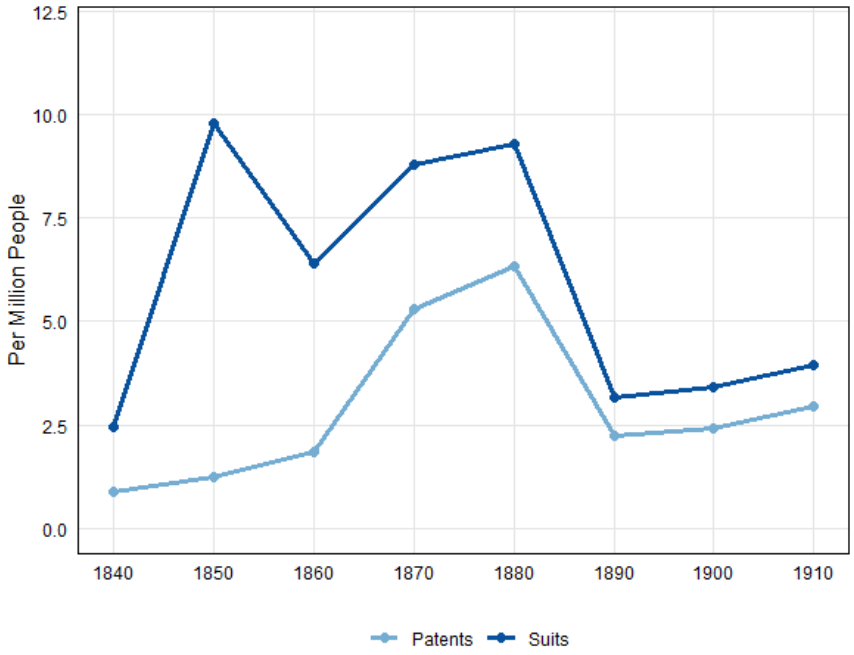
[36] Again, Katznelson's estimate accords remarkably closely with this paper's estimate, which finds that the average reporting rate for the pre-war years (1937–1941) was ~86.5%.

[37] A legal change in 2011 altered the joinder rules for patent suits.

reports—the latest litigation wave is very familiar, but observers of the patent system in the early twentieth century also remarked on surges of patent litigation involving new technologies, such as radio.[38]

For the nineteenth century, the leading source of information about the volume of patent litigation is Beauchamp (2016).[39] Unfortunately, before 1923, there is no centralized repository of patent litigation statistics—records are held in the archives of individual trial courts. Beauchamp hand-reviews records from two of the busiest trial courts of the nineteenth century—the Circuit Courts for the Southern District of New York (SDNY) and the Eastern District of Pennsylvania (ED Pa)—on one-decade intervals between 1840 and 1910, and reports the number of suits and the number of (unique) patents asserted in litigation, which are shown below in Figure 2.

Figure 2: Patent Litigation 1840–1910, SDNY & ED Pa



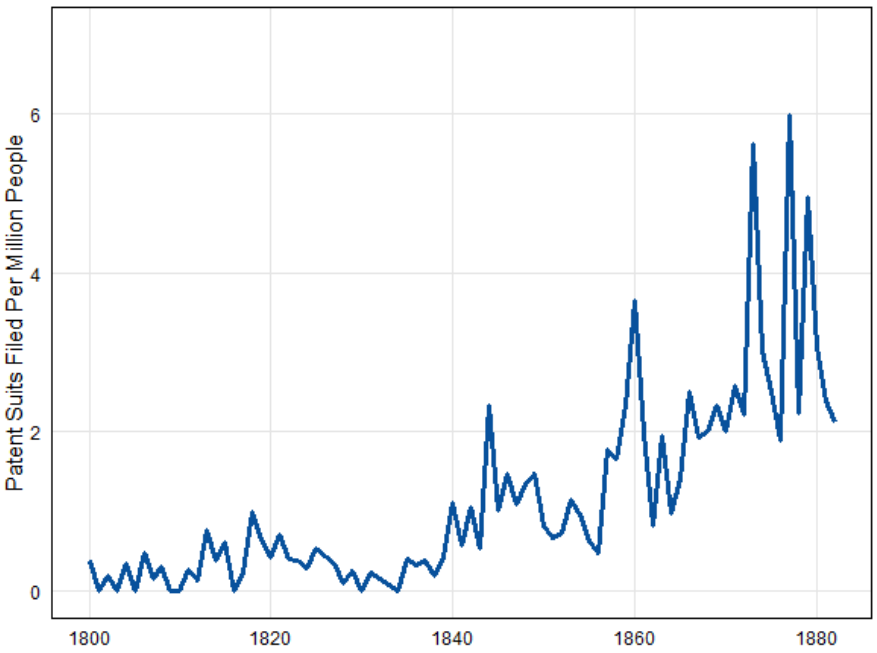
These data show a large wave of patent litigation in the second half of the nineteenth century. The center of gravity of this wave seems to be in the 1870s and 1880s, although the year 1850 also

[38] *See, e.g.*, Katznelson, *supra* note 1, at 9.
[39] Beauchamp, *supra* note 1.

shows a large spike of patent suits (filed almost entirely by a single patent owner).[40]

These results also match a second source of nineteenth century data. Working with the Boston Public Library, the National Archives at Waltham, and the Internet Archive, Rosen compiles a database of patent suits filed in the Circuit Court for the District of Massachusetts (D Mass) from 1800–1882,[41] quite likely the second-busiest patent-litigation trial court of the nineteenth century after SDNY. The results are shown below, scaled per capita, in Figure 3.

Figure 3: Patent Litigation 1880–1882, D Mass



These data similarly show a large wave of patent litigation in the late nineteenth century, peaking in the 1870s and 1880s. They also have two other noteworthy features. First, the yearly data show the high variability of patent filings from year to year (so sampling on low frequencies could pick up outliers). Second, these data do not show any large, sustained volume of litigation around

[40] *Id.*
[41] Zvi Rosen, *Patent Case Files from the Archives: New England Edition* (July 12, 2017), available at https://mostlyiphistory.com/2017/07/12/patent-case-files-from-the-archives-new-england-edition/; *see also* Zvi Rosen, https://x.com/zvisrosen/status/1336387798862016523.

1850. Together, these features are additional evidence that the center of gravity of the nineteenth century litigation wave was likely in the 1870s and 1880s.[42]

Lastly, the story told by these datasets also matches reports from contemporary observers in the late nineteenth century, who remarked on the large volume of patent litigation passing through the courts.[43]

In sum, there is evidence that there have been three large surges of patent litigation in U.S. history—one in the 1990s and 2000s, another in the 1920s and 1930s, and a third in the 1870s and 1880s.

III. DATABASE

This Part describes the litigation database that the paper constructs. Part A covers the transcription of the Gazettes, and Part B covers the assembly of the final database.

---

[42] At least a little caution should be taken in interpreting these figures. SDNY, ED Pa, and D Mass were extremely busy patent-litigation venues and likely capture a substantial fraction of patent cases in these years (perhaps ~50% based on published opinions), but they may not be representative. It is also possible that observed changes in these venues reflect changes in the *distribution* of litigation, rather than the *volume*. For example, perhaps eastern venues dominated for much of the nineteenth century, while western venues (like the Northern District of Illinois (and eventually the Southern District of California) became more important over time.

Even more importantly, as Beauchamp notes, great caution must be taken in comparing litigation volumes across the nineteenth and twentieth centuries. An important clue about why is visible in the SDNY / ED Pa data—the changing ratio of suits to unique patents over time. In the early years, this ratio was quite high. For example, in 1860, there were ~3.5 suits for each patent. By 1900, the ratio had fallen to ~1.4. And the Gazette data shows that this ratio continued to fall in the early twentieth century, eventually stabilizing around ~.8 in the 1930s, where it remained for much of the rest of the twentieth century.

This likely reflects the changing concentration of economic activity, which affects how patent infringement allegations are packaged into lawsuits. In the early nineteenth century, the U.S. economy was composed of many small artisans. In this world, to capture all sales of an infringing product, a plaintiff needed to file many separate lawsuits asserting the same patent against many small producers (which will produce a high ratio of lawsuits to unique patents). In the late nineteenth and early twentieth centuries, economic concentration rose—small artisans were replaced by larger corporations. In this new world, a plaintiff could achieve the same coverage by suing the smaller number of large producers. This will mechanically reduce the volume of lawsuits, as well as the ratio of suits to patents (exactly what we see). Beauchamp also reports other evidence consistent with this story—over the late nineteenth and early twentieth centuries, the parties in patent cases shift from individuals to corporations. *See* Christopher Beauchamp, *Dousing the Fires of Patent Litigation*, in THE BATTLE OVER PATENTS 136 (Stephen H. Haber & Naomi R. Lamoreaux eds., 2001).

[43] Beauchamp, *supra* note 1.

A. Transcribing the Gazettes

The key tool to construct the litigation database is a February 1922 amendment to the patent statute that requires clerks of court to send a notice to the Patent Office each time a case arising under the patent laws is filed or decided.[44]

By law, the notices must report the names of the parties and the numbers of the asserted patents. But, as a matter of practice, they also contain the docket number of the case, the court in which it was filed, and the date of the event that initiated the notice. The reporting provision was later re-codified in the major 1952 amendments to the patent statute, where it remains.[45]

Historically, the Patent Office published these notices in weekly issues of the Official Gazette, which is the official publication of the Patent Office. The first list of patent suits was published in the July 4, 1922 issue, and the last list was published on December 12, 1984. An example list is reproduced on the following page.

A key challenge in capturing the information in these lists is the volume. The Gazettes contain roughly 75,000 individual litigation notices, spread across roughly 3,000 issues over 60 years. It would likely take a human many thousands of hours to record this information manually.

To handle the volume, the paper turns to a large language model (LLM).[46] LLMs are well-suited for this task—they are able to understand the content of the notices, find key details for each suit, and record them in a structured format. And, critically, an LLM can do all of this *much* faster than a human.

At a high level, the process was the following. A high-quality PDF scan of each Gazette issue was obtained. The list of patent suits was then extracted and fed to an LLM alongside a standard prompt that asked the LLM to find each suit on the list and capture key details. Finally, the outputs for each Gazette issue were combined together into a dataset spanning all Gazettes. Human supervision was done throughout.

---

[44] R.S. § 4921 (1922).
[45] 35 U.S.C. § 290.
[46] In particular, Claude 3.7 Sonnet was used. Relative to other LLMs, Claude has three desirable properties: (1) an excellent PDF reader; (2) willingness to process an entire Gazette issue without repeatedly stopping to ask for additional feedback; and (3) model stability over the relevant period (February to May 2025).

Figure 4: Official Gazette January 21, 1930

years. It made up this word from the words "Do" and "more" just as plaintiff had done. It used capital letters as in the drawing in plaintiff's application for the mark. Before it went into production it was given notice of the strong probability of confusion in the trade due to its adoption of the word. After production began, it was again notified, but persisted in its use. It designed and used a label for its products in which the word "Dumore" is given prominence. On the label is printed the words "The Dumore Company, Dayton, Ohio, U. S. A.," but this name and address inconspicuous as compared with the word "Dumore." Its product was electrically driven by small motors with lamp-socket attachment just as were plaintiff's products. It was, of course, for domestic use, just as were many of plaintiff's appliances. Products similar to those of plaintiff are often found in the same stores with electric washing machines. Defendant sold about 400 of its machines up to November, 1927. It is true that the vice president of defendant testifies that he had never heard of the name "Dumore" as a trademark before he adopted it but the record sheds no light as to what any other officer or agent of the company knew. If he had never heard of it, it is a strange coincidence, not only that he selected plaintiff's word, but that he made it up from the same words in the same form and misspelled in the same way. It was not essential that plaintiff prove any particular injury or that any purchaser was misled. *Bickmore Gall Cure Co. v. Karns*, 134 F. 833, 835 (C. C. A. 3). It is sufficient if there was manifest liability to deceive, *Ralston Purina Co.* v. *Western Grain Co.* 23 F. (2d) 253, 255 (C. C. A. 5), or, if there was reasonable probability of injury, *Peninsular Chem. Co.* v. *Levinson*, supra. That there was such reasonable probability is obvious. Illustrations may be found in *Vogue Co.* v. *Thompson-Hudson Co.*, supra; *Peninsular Chemical Co.* v. *Levinson*, supra, and *Wall* v. *Rolls-Royce*, supra.

That plaintiff is entitled to relief is clear. The relief granted by the district court was incidental to the jurisdiction it assumed in determining the validity of the word "Dumore" as a technical trademark. We conclude that it should be based upon the theory herein set forth. The district court enjoined the defendant from the further use of the name "Dumore" as a trade-mark on its washing machines in connection with the corporate name, "the Dumore Company," or any other corporate name or business title which includes the word "Dumore." The effect is to allow defendant to use the word "Dunmore" alone. But the word, in substantially the same form, has been so long associated in the trade with plaintiff's goods that there is a likelihood shading almost into certainty that any electrically-driven device labeled "Dumore" only would be regarded as plaintiff's product. To obviate such manifest probability, the defendant may use the word "Dumore" only when accompanied with its present corporate name, Dayton Washer Company. It may not use it otherwise.

Thus modified, the decree is affirmed.

Patent Suits

[Notices under sec. 4921, R. S., as amended Feb. 18, 1922]

1,090,370, Re. 13,932, 1,105,053, Wylde & Schenck, Door-controlling device, appeal filed Sept. 6, 1929, C. C. A., 3d Cir., Doc. 4251, *Elevator Supplies Co.* v. *Graham & Norton Co.*

1,092,078, Re. 13,626, M. C. Overman, Vehicle tires, appeal filed Nov. 26, 1929, C. C. A., 2d Cir., Doc. 10699, *Overman Cushion Tire Co., Inc.*, v. *Goodyear Tire & Rubber Co., Inc.*

1,102,130, G. S. Bennett, Vacuum cleaner, filed Aug. 10, 1929, D. C., E. D. N. Y., Doc. 4386, *J. J. Duffie* v. *Brooklyn Edison Co., Inc.*

1,105,053. (See 1,090,370.)

1,125,476, G. Claude, System of illuminating by luminescent tubes, filed Aug. 10, 1929, D. C., E. D. N. Y., Doc. 4385, *Claude Neon Lights, Inc.*, v. *Hy-Glo Tube Lights, Inc., et al.* Doc. E 3159, *Claude Neon Lights, Inc., et al.* v. *Cenco Neon Laboratories, Inc., et al.* Consent decree for plaintiff Nov. 20, 1929.

1,132,657, R. B. Gilchrist, Ice-cream ladle, D. C., S. D. N. Y., Doc. E 45/164, *The Gilchrist Co.* v. *Guaranteed Ice Cream Disher Co. et al.* Consent decree for plaintiff Nov. 23, 1929.

1,140,469, A. P. Lundin, Boat, filed Sept. 13, 1929, D. C., E. D. N. Y., Doc. 4427, *Welin Davit & Boat Corp.* v. *C. M. Lane Life Boat Co., Inc.*

1,174,403, F. K. Fassett, Door-operating mechanism; 1,565,143, H. J. Fear, Elevator-door control; 1,587,007, C. Norton et al., Door-operating mechanism for elevators, appeal filed Sept. 6, 1929, C. C. A., 3d Cir., Doc. 4250, *Elevator Supplies Co.* v. *Graham & Norton Co.*

1,210,492, 1,302,057, J. G. Knight, Outlet box, filed Oct. 4, 1929, D. C., E. D. N. Y., Doc. 4463, *J. G. Knight* v. *Tollner Electric Co., Inc.*

1,219,560, J. G. K'ng, Engine starter; 1,525,923, A. A. Densmore, same; 1,483,872, same, Power-transmission mechanism, filed Sept. 30, 1929, D. C., E. D. N. Y., Doc. 4451, *A. A. Densmore* v. *L. A. D. Motors Corp.*

1,262,860, S. B. Smith, Incubator, filed Nov. 22, 1929, D. C., S. D. Calif. (Los Angeles), Doc. Q-69-M, *S. B. Smith et al.* v. *H. H. Washam.*

1,271,529, M. C. Hopkins, Acoustic device, filed Sept. 26, 1929, D. C., E. D. N. Y., Doc. 4450, *Lektophone Corp.* v. *Colonial Radio Corp.*

1,277,713, L. Goldstone, Boot or shoe counter reinforce, D. C., W. D. N. Y., Doc. 1307-G, *O. Schindler* v. *A. G. Stevenson & Co., Inc.* Dismissed Nov. 12, 1929.

1,285,593, Baumann & Butz, Round-cornering machine for book binding, filed Nov. 14, 1929, D. C., E. D. Mo. (St. Louis), Doc. 8835, *C. P. Martin* v. *United Shoe Machinery Corp.*

1,302,057. (See 1,210,492.) 1,307,733. (See 1,475,980.)

1,307,734. (See 1,475,980.)

1,334,787, O. F. Palmer, Machine for making metallic conduits, filed Oct. 8, 1929, D. C., E. D. N. Y., Doc. 4476, *Triangle Conduit Co., Inc.*, v. *Federal Armored Cable Co., Inc.*

1,379,224, O. P. Smith, Dog-racing amusement, filed Sept. 5, 1929, D. C., E. D. N. Y., Doc. 4414, *H. M. Smith* v. *E. Roberts et al.*

1,388,546, J. Brueck, Nail polish, filed Aug. 27, 1929, D. C., E. D. N. Y., Doc. 4395, *Patent Owners, Inc.*, v. *H. Marks et al.*

1,415,213, R. W. White, Mold for casing clay articles, filed Oct. 16, 1929, D. C., E. D. N. Y., Doc. 4496, *H. W. Dyer* v. *Alexander Oil Burner Corp.*

1,474,461, E. A. Bircher, Fermentation seal, filed Oct. 15, 1929, D. C., E. D. N. Y., Doc. 4495, *E. A. Bircher* v. *Consumer's Product Co., Inc.*

1,475,980, O. Zerk, Lubricating apparatus; 1,307,733, A. B. Gullborg, same; 1,307,734, same, Lubricating means, D. C., E. D. N. Y., Doc. 4410, *Alemite Corp.* v. *N. L. Sper.* Decree for plaintiff Nov. 21, 1929.

More specifically, high-quality PDF scans of each weekly Gazette issue between July 1922 and December 1984 were downloaded from Google Books (using indices compiled by HathiTrust), and the list of patent suits in each Gazette was extracted. Initially, the lists from the first six months of Gazette issues in each decade (1930–1980, plus 1923) were manually reviewed to understand the structure of the data. A standard prompt was then created through trial-and-error over several years of Gazette data until it produced consistently excellent results. The full text of the prompt is reproduced in Appendix A.

Each list of suits was then fed to the LLM, alongside the prompt. The LLM read the list and produced a table with one row per suit and columns for each piece of information. The output table was manually reviewed for errors if any of the following six conditions were met:

1. Missing Information. Each output table should typically should have information for each non-date field (docket number, plaintiff, defendant, court, and patents) plus at least one date field (filing, decision, or notice). Suits with missing fields were reviewed.

2. Out of Order. Within each Gazette issue, the lists of suits are (invariably) organized in order of the lead patent. The rows of the output table therefore also appear in lead-patent order, if the LLM processes them sequentially. If the rows are instead out-of-order, that is an important signal of an OCR issue. The pages of the Gazettes are split into two columns, and, occasionally, the OCR does not correctly parse them, and instead jumps back and forth across the columns. Typically, the LLM was able to correctly handle this issue, but occasionally it leads to errors. Out-of-order suits in the output table are a signal of this problem.

3. Patent Column Limit. The prompt instructs the LLM to create columns for up to 20 patents per suit. Of course, some suits (rarely) have more than 20 patents. But it is not feasible to specify the correct number of patent columns in advance (the largest number of patents associated with a single suit is known only after the data are processed). Creating large numbers of columns is cumbersome. And it does not work to tell the LLM to simply create as many patent columns as necessary for a particular Gazette (LLMs cannot count very well). The better approach proved to be telling the LLM to create 20 patent columns by default, and manually reviewing any suits that hit the limit (to check for extra patents).

4. Long Lists. The LLM reliably found all suits on almost all Gazette lists. But it had a slightly higher tendency to skip suits on the very longest Gazette lists, with 80+ suits. In particular, this tendency manifested at the starts and ends of columns (where the LLM had to "jump" to continue). Accordingly, manual checks were done of the starts and ends of columns on very long lists.

5. Last Patent. Some Gazette lists span a single page, while others span multiple pages. Occasionally, on a multi-page list, the LLM would forget to look at the last page. To catch this issue, manual review was done every time the LLM's output table did not end on the last patent in the Gazette list.

6. Other. Manual review was also done for any other clear error, plus occasional spot checks to make sure performance was running well over time.

Finally, once all Gazette lists were processed, a post-processing review was carried out to check for a variety of special, hard cases. In particular, every single Gazette PDF in the set was manually opened a second time and checked for the following six categories of issues:

1. Errata. The Gazettes occasionally contain errors (e.g., a misprinted patent number). When the Patent Office catches an error, it will print an erratum in a later Gazette issue. There is no natural way for the LLM to handle this issue. Instead, it was addressed by manual review.

2. Marks, Blemishes, etc. Scan quality was generally quite high. But some scans had issues that could impair legibility, such as an ink spot, smudge, or cut-off page. Each Gazette scan was checked for these issues, and the LLM output was manually reviewed where they arose.

3. **Date Issues (1922–1926)**. The Gazette issues in the early years (1922–1926) have a collection of features that, together, made it more difficult for an LLM to correctly classify dates. For example, these Gazettes report that a case was "filed," but also often say that an order deciding a case was "filed" on the docket. Starting in mid-1926, the Gazettes clean up their date reporting, which improved LLM date classification accuracy to ~perfection. A manual review of all dates in the early Gazettes was done to catch errors.

4. Transfers. In some cases, Gazette entries contain a notification that a suit has been transferred from one jurisdiction to another. Often, later Gazette issues then contain notifications about events in the transferred suit,

which creates a risk of duplicate entries. These notes were manually recorded, and resolved by keeping the original suit.

5. OCR Layer. As noted above, the Gazette column structure sometimes introduced OCR layer issues. Generally, the LLM handled these correctly, and, even when not, these issues generally manifested as an out-or-order output table. However, it is possible that an OCR layer issue could create an unnoticed problem (where the LLM skips some suits but otherwise remains in-order). This is a particularly dangerous issue because OCR-layer problems are not constant over time; rather, a pair of formatting changes in the early 1950s—eliminating the vertical line between columns, and moving the two columns closer together—substantially increased their prevalence. Accordingly, failing to catch OCR issues might introduce variation over time. This was handled by manually checking every scan for OCR layer issues and reviewing the output.

6. Scan Resolution. In the mid-1950s, the Gazettes also began bolding patent numbers, which can impair the legibility of numbers in low-resolution scans. Despite this issue, LLM performance remained remarkably good. For example, the next page shows perhaps the single worst scan in the entire database (from August 1956). To the human eye, many of the patent numbers look like little more than inkblots. Nevertheless, the LLM achieved 100% accuracy on these patents (it is possible to check by using the listed inventor/invention information). However, occasional errors did occur; particularly, there was a small tendency to confuse 3s and 8s in bad scans. This is also another dangerous issue that could introduce variation in transcription accuracy over time (as bolding began in the early 1950s). To address it, every scan was manually reviewed for quality, and patent numbers were checked on low-quality scans.

It is essential that the Gazettes are accurately transcribed. To check, one full weekly Gazette issue was randomly drawn from each of the 63 years in the dataset—1922 to 1984—and accuracy was hand-checked on the three most important and sensitive pieces of information for every suit: (1) whether each listed suit was captured (or some were missed); (2) filing dates, and (3) patent numbers.

Figure 5: Official Gazette August 14, 1956

ent 2,247,188 held invalid and not infringed; counterclaims dismissed July 12, 1956.

2,253,900, J. H. Speed, Tool handle, filed July 13, 1956, D. C. Oreg. (Portland), Doc. 8704, *Louise K. Webster* v. *Speed Corp.*

2,247,188. (See 2,190,442(b).)

2,278,437, P. H. Gentsel, Safety valve construction; **Re. 22,164**, same, Valve mechanism, filed Nov. 10, 1950, D. C., S. D. N. Y., Doc. 48/338, *Foster Engineering Co. et al.* v. *Manning, Maxwell & Moore, Inc.* Interlocutory judgment for injunction; case referred to Special Master Apr. 18, 1955; C. C. A. reversed judgment of District Court; complaint dismissed Mar. 19, 1956; C. C. A. judgment made judgment of District Court July 9, 1956.

2,305,261, M. M. Kinley, Method of removing pipe from wells, filed Apr. 20, 1955, D. C., S. D. Tex. (Houston), Doc. 8877, *Ford I. Alexander* v. *Jack Foster Wire Line Service, Inc. et al.* Consent decree; patent held valid; claims 1 and 2 of patent held infringed; injunction granted July 2, 1956.

2,377,349. (See 2,190,442(a) and (b).)

2,412,262, Maxwell and Edwards, Metallic clip for connecting and reinforcing joints in wood structures, filed July 9, 1956, D. C., S. D. Calif. (Los Angeles), Doc. 20164, *Homer B. Maxwell et al.* v. *Newton Products Co. et al.*

2,452,810. (See 2,722,256.)

2,459,110, S. P. Midouhas, Collapsible clothesline, filed July 16, 1956, D. C., E. D. N. Y. (Brooklyn), Doc. 16724, *Kamkap, Inc.* v. *J F D Mfg. Co., Inc.*

2,476,728, F. Klumpp, Jr., Solderless blade for plug caps, filed July 9, 1956, D. C., E. D. N. Y. (Brooklyn), Doc. 16705, *Heyman Mfg. Co.* v. *Lectroloid Corp.*

2,515,817, K. C. Augenstein, Expansion bracelet; **Des. 165,355**, same, Expansible chain for a bracelet or similar article, filed Oct. 7, 1954, D. C., E. D. N. Y. (Brooklyn), Doc. 14834, *Speidel Corp.* v. *Glen Corp.* Consent judgment July 10, 1956.

2,537,980, P. J. Warmath, Tag holder and protector, filed Jan. 18, 1956, D. C., S. D. Calif. (Los Angeles), Doc. 19347–C, *Pit-All Pricing Corp.* v. *Monadnock Mills, Inc. et al.* Patent held valid; defendant Mankot, Inc. has infringed; defendant Monadnock Mills threatened infringement; defendants restrained from further infringement (notice July 3, 1956).

2,546,466, 2,712,419, F. J. Martini, Spinning type fishing reel, filed July 12, 1956, D. C., E. D. Mich. (Detroit), Doc. 15830, *Franz J. Martino* v. *Great Lakes Products, Inc.*

2,571,550. (See Re. 23,694.)

2,586,211, S. B. Beliaeff, Toy motor, filed July 3, 1956, D. C. N. J. (Newark), Doc. 538/56, *Stephen B. Beliaeff et al.* v. *Marex Mfg. Co., Inc. et al.*

2,646,143, Daleo and Foglio, Sample display case, filed July 1, 1956, D. C., S. D. N. Y., Doc. 111/21, *Graphic Arts Center Inc.* v. *Rapid Art Service Inc.*

2,651,941, L. Sunderland et al., Overcoat, filed July 13, 1956, D. C. Mass. (Boston), Doc. 56/591–F, *Louis Sunderland Co.* v. *Lebow Brothers, Inc.*

2,668,328, F. E. Porter, Method of casting patterned plastic sheets, filed Mar. 18, 1954, D. C. N. J. (Newark), Doc. 193/54, *U. S. Chemical Corp.* v. *Plastic Glass Corp.* Patent held invalid; order of dismissal July 2, 1956.

2,668,206, F. Kern, Jr., Gas regulator, filed July 11, 1956, D. C., E. D. Mich. (Detroit), Doc. 15824, *Maxitrol Co.* v. *General Controls Co.*

2,693,789, L. J. Lechtenberg, Air-cooled internal-combustion engine, filed July 13, 1956, D. C., E. D. Wis. (Milwaukee), Doc. 6/c/136, *Briggs & Stratton Corp.* v. *Tecumseh Products Co.*

2,693,926, J. Tatko, Pallet, platform or the like, filed July 6, 1956, D. C. Maine (Portland), Doc. 4/191, *Tatko Brother Slate Co., Inc.* v. *L. C. Andrew.*

2,694,580, H. Head, Composite wood and metal ski having plastic running surface, filed July 13, 1956, D. C. Md. (Baltimore), Doc. 8998, *Head Ski Co., Inc.* v. *Kam Ski Co., Inc. et al.*

2,704,211, C. Decepoli, Shuffleboard weight, filed July 10, 1956, D. C. N. J. (Newark), Doc. 546/56, *Carmine Decepoli* v. *National Shuffleboard Co.*

2,712,419. (See 2,546,465.)

2,712,429. (See Des. 176,657.)

2,720,277, C. E. Meyerhoefer, Filter structures for vacuum cleaners, filed July 13, 1956, D. C., E. D. N. Y. (Brooklyn), Doc. 16723, *Lewyt Corp.* v. *Edward J. Poggi.*

2,722,256, J. E. Hise, Quartering machine for potatoes and the like; **2,453,819**, L. V. Uglow, Seed potato cutting machine filed July 13, 1956, D. C. Nebr. (Omaha), Doc. 0190, *James Edward Hise* v. *Lockwood Grader Corp. et al.*

2,722,712, M. Yellen, chair structure, filed July 3, 1956, D. C. S. D. N. Y., Doc. 111/43, *Yellen Inc. et al.* v. *Frank & Son, Inc.*

2,728,947, 2,728,948, J. L. Kallus, Molds for making dolls composed of plastic material; **2,728,980**, same, Method for producing molds for making dolls composed of plastic material, filed July 3, 1956, D. C., S. D. N. Y., Doc. 111/22, *Whitkall Corp. et al.* v. *Perfect Doll Moulds Co., Inc.*

2,728,948. (See 2,728,947.)

2,728,980. (See 2,728,947.)

2,732,327, W. K. Teller, Anti-perspirant stick, filed June 29, 1956, D. C. Del. (Wilmington), Doc. 1832, *Northam Warren Corp.* v. *Pharma-Craft Corp.*

2,743,978, C. Shore, Leg and brace construction for folding table with drop leaves, filed June 29, 1956, D. C., E. D. Pa. (Philadelphia), Doc. 20959, *Falco Products Co.* v. *Lit Brothers, Division of City Stores Co.* Same, filed July 3, 1956, D. C., E. D. N. Y. (Brooklyn), Doc. 16691, *Falco Products Co.* v. *Gertz, B. Inc.*

2,749,725, Essman and Lavern, Portable air conditioning apparatus, filed July 5, 1956, D. C., N. D. Ill. (Chicago), Doc. 56c1141, *The Eska Co., Inc.* v. *Parker Metal Products.*

Re. 22,164. (See 2,278,437.)

Re. 23,186, W. H. Mead, Surface treating method and apparatus, filed July 12, 1956, D. C., N. D. Calif. (San Francisco), Doc. 35667, *Educt-O-Matic Corp.* v. *Vacu-Blast Co., Inc. et al.*

Re. 23,694 (of 2,571,550), L. G. Ehmann, Material handling device for industrial trucks, filed July 12, 1956, D. C., N. D. Ill. (Chicago), Doc. 56c1180, *Hyster Co.* v. *Hunt Foods, Inc.*

Des. 164,922, B. D. Gould, Textile fabric, filed Feb. 11, 1953, D. C., S. D. N. Y., Doc. 82/314, *Trojan Textile Corp. and Bernard D. Gould* v. *Crown Fabrics Corp.* Complaint amended to add Bernard D. Gould as coparty plaintiff Nov. 3, 1955; Design patent held invalid (notice July 12, 1956).

Des. 165,355. (See 2,515,817.)

Des. 168,143, M. Cohen, Garment display rack, filed July 9, 1956, D. C., N. D. Tex. (Dallas), Doc. 6579, *Sel-O-Rak Corp.* v. *Revolving Display Mfg. Co. et al.*

Des. 172,923, H. Federman, Watch, filed Nov. 10, 1954, D. C., S. D. N. Y., Doc. 96/358, *Henry Federman* v. *Enicar Watch Corp. et al.* Order of discontinuance (notice June 26, 1956).

Des. 176,657, H. Dreyfuss, Control instrument; **2,712,429**, W. A. Ray, Fluid control valve (Count I as to Des. 176,657, and Count II for a Declaratory Decree as to Patent 2,712,429), filed July 11, 1956, D. C., N. D. Ill. (Chicago), Doc. 56c1173, *Minneapolis-Honeywell Regulator Co.* v. *General Controls Co.*

The drawing process yielded a relatively large sample. In total, these Gazettes contain 2,120 suits, 1,742 filing dates, and 3,401 patents.[47] Results are shown in Table 1 below.

Table 1: Transcription Accuracy

Suits	
Present	2,120
Missed	1
Accuracy	99.95%
Filing Dates	
Present	1,742
Missed	0
Misclassified	1
Accuracy	99.94%
Patents	
Present	3,401
Missed	12
Number Error	1
Matched to Wrong Suit	1
Accuracy	99.59%

Overall accuracy was high across all categories. The nature and sources of the errors are discussed below.

1. Suits. Here there was a single error—the LLM skipped over one suit at the bottom right of a column in a Gazette issue (the beginnings and ends of columns can be a tricky area for an LLM).

2. Filing Dates. The error here was misclassification—the LLM wrongly classified a decision date as a filing date in a 1923 Gazette issue. The challenges with date classification in early-year Gazettes were discussed above. Manual review was done to identify and correct date errors, but this error was missed.

[47] There are fewer filing dates than suits because not all suits have a filing date (some instead have a reported decision date), and there are more patents than suits because some suits assert more than one patent.

3. Patents. There were several errors. First, the LLM missed 12 patents. One missed patent came from the missed suit noted above (which asserted a single patent). Three others were simply missed entries on longer lists (e.g., a suit involved four patents, and the LLM spotted three). The remaining eight missed patents came from two errors. In one case, the list of patents asserted in a suit stretched across two pages, and the LLM missed four patents on the first page. In the other case, the LLM seems to have misread a complex decision—the court entered a decision dismissing four patents from the case, and the LLM seems to have interpreted the decision as meaning that the patents should be treated as never having been asserted at all.[48]

Finally, two other patent errors occurred. One was a numerical error—the LLM misread a 3 in a low-quality scan as an 8. As noted above, manual review was done to correct these errors, but this one was missed. In the other case, the LLM made a mistake parsing two long lists of patents asserted in two related cases—it mistakenly put a patent from the first list on the second list.

In sum, despite several errors, transcription accuracy was at least 99.5% in all categories. In the author's experience, a key reason why the process seemed to work well is that LLMs and humans have complementary skills (and weaknesses) on this task. In other words, humans tend to make mistakes that are easy for an LLM to avoid, and LLMs tend to make mistakes that are easy for a human to catch. For example (as the author learned), a human digitizing these records by hand will tend to make typos—something the LLM almost never does. And the LLM is able to go blazingly quickly. By contrast, an LLM processing a multi-page list of patent suits will occasionally[49] forget to look at the second page—something that is easy for a human to catch (but requires supervision). And there are many other examples of this character,

---

[48] In this particular instance, thoughtful behavior by the LLM (trying to understand the meaning of a decision) did have an adverse consequence. But, in general, the LLM had no difficulty parsing decisions. And this kind of thoughtfulness had many benefits elsewhere that more than outweighed its occasional costs.

For example, in one Gazette issue, a patent was missing a digit due to a misprint. The Gazette lists often contain various cross-references that effectively have extra copies of the patent numbers. The LLM noticed that a digit was missing (presumably because it understands what patents should look like), located the appropriate cross-reference, and correctly plugged in the missing digit, all without any explicit instruction—the LLM was simply thoughtfully carrying out its mandate in a special case.

[49] Less than one time in a hundred, but more than one time in a thousand.

i.e., hard cases for an LLM that can be resolved with human review. Complementary skills are the perfect case for collaboration—working together, an LLM and a human can be much faster and more accurate than either alone.

B. Compiling the Database

Once the Gazettes were transcribed and assembled, there were two further steps to create the final database.

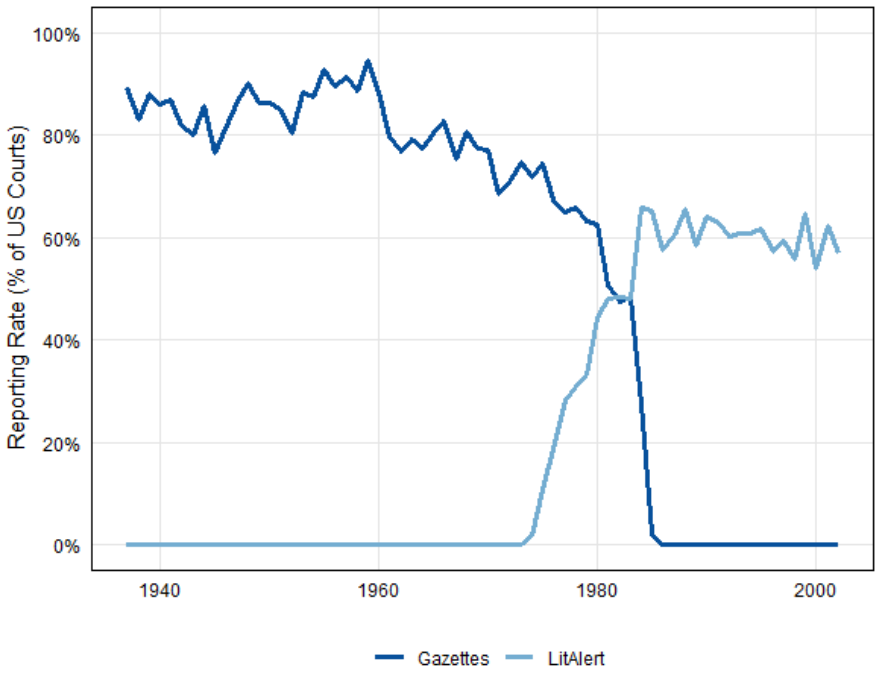
One was combining the Gazette data and the LitAlert data. Recall that the published Gazettes cover the period 1923–1984, while the commercial database LitAlert picks up coverage of the notices post-1984. However, LitAlert also has some coverage in the early 1980s, so there are some years of overlap where we want to select the strongest source.

The second (and related) issue was measuring and accounting for under-reporting. The 1922 statutory amendment requires clerks of court to send litigation notices to the Patent Office. However, that does not mean they always do. Some clerks might, for example, be unaware of their statutory obligation, or occasionally forget to send the notices.

To begin, it is helpful to look at the reporting rates in each dataset, using the complete U.S. Courts data as a baseline. This shows where each dataset has stronger and weaker coverage, as well as overall variation in reporting. Figure 6 on the following page shows the reporting rates for the Gazettes and LitAlert in each fiscal year, 1937–2002.

The Gazettes have relatively strong coverage in the early years, from 1923–1980, and gradually fall off in the early 1980s. This makes sense—the Gazettes stopped publishing litigation notices altogether after 1984, and even before the cutoff the Gazettes will miss, e.g., a suit filed in 1984 that was not reported to the Patent Office until 1985. LitAlert has spotty coverage in the 1970s but becomes more comprehensive over time, reaching a steady state of ~60% in 1984. Lastly, Both datasets have somewhat weaker coverage in the early 1980s (1981–1983). In these years, the Gazettes begin to lose cases, but LitAlert is still ramping up. However, at the margin, the Gazettes are still slightly strongly in 1981, while LitAlert becomes stronger in 1982.

Figure 6: Reporting Rates



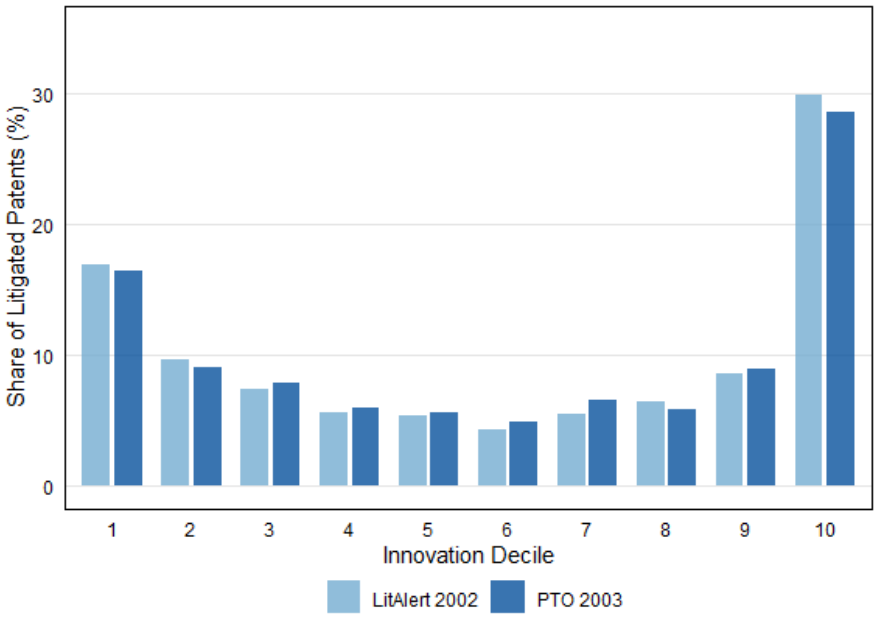
Accordingly, the final database will join the datasets along that line, using the Gazettes for 1923–1981 and LitAlert for 1982–2002. In total, the combined database has ~60,000 patent suits and ~90,000 asserted patents. Excluding design patents and plant patents—two special types of patents that are not scored by the Kelly et al. (2021) metric—the database has ~55,000 suits and ~85,000 asserted patents.

Still, Figure 6 also shows that, even after selecting the years with the strongest coverage, the database will have some under-reporting, whose magnitude has varied over time. The reporting rate is strong in the early years (~85%), but decays to ~60% by the end of the 1970s, where it remains on average thereafter (with a drop to ~50% in the early 1980s).

A key question is therefore whether the pool of reported cases differs substantially from the broader pool of all filed cases. We can get at least some insight into this issue. The Patent Office releases essentially complete patent-level litigation data starting in 2003, and the pool of litigated cases tends to change relatively gradually from year-to-year. Accordingly, we can compare the reported sample in 2002 (from LitAlert) to the full sample in 2003 (from the Patent Office) to see whether the two pools of cases look relatively similar, or whether there is a large discontinuity or step-

change that would suggest a large reporting bias. Looking at 2002/2003 is also a good window into the most serious under-reporting concern, which is that lower reporting rates in recent years will bias coverage of the second major technological wave. Figure 7 shows the comparison between the LitAlert and Patent Office datasets.

Figure 7: Patent Office v. LitAlert



Overall, the two pools of cases are quite similar—the distributions show similar numbers of patents across all deciles of the innovation metric. There is a small amount of year-to-year variation, but no large step change that would indicate a major reporting bias or change the basic picture of the pool of litigated patents. Of course, a problem cannot be ruled out completely; it is certainly possible that the nature of the reporting bias has changed over time. But this is at least some reassurance that reporting bias is unlikely to be a large, first-order error term in the results.

A final issue is accounting for variation in the reporting rate over time. We can do this by weighting the data in each year by the reporting rate.[50]

[50] There are two special cases. First, before 1937, we do not have U.S. Courts data. For these years, the paper will use the average reporting rate in the pre-war years (1937-1941), which is ~86.5%. Second, the U.S. Courts data for 1992 runs from October 1 to September 30 (rather than July 1 to June 30), which

IV. RESULTS

This Part reports two results. First, technological revolutions have played a large role in explaining waves of patent litigation. Second, a key feature of top-scoring, revolutionary patents that drives litigation is their unusual longevity.

A. Patent Litigation Waves

Recall, the Kelly et al. (2021) measure scores each issued U.S. patent from 1840–2002 based on two criteria: (1) textual novelty (over past patents) and (2) textual influence (over future patents).[51] The authors use these scores to construct an innovation index by looking at the number of patents in the highest decile (top 10%) issued by the Patent Office each year.[52] Waves of top-scoring patents reveal major waves of basic innovation. Figure 8 on the following page shows this index for 1840–2002.

The innovation index reveals three major technological waves. There is a first wave in late nineteenth century (or perhaps two close together), featuring advances in manufacturing, railroads, and electricity. Top-scoring patents from this wave include patents on textiles and sewing machines, electric light, railroad improvements (such as the air brake), and concrete.[53] There is a second wave in the 1920s and 1930s, featuring advances in chemicals, electronics, and automobiles. Top-scoring patents from this wave include patents on bakelite (the first synthetic plastic), nylon, radio antennas, X-ray tubes, and engine ignition.[54] Finally, there is a third wave in the late twentieth and early twenty-first centuries, featuring advances in computers, telecommunications, and biotechnology. Influential patents from this wave include patents on polymerase chain reaction methods, transgenic mice,

---

introduces an inconsistency in the periods. For 1992, the paper will use the average reporting rate for the modern period (1984–2002), which is ~60.7%.

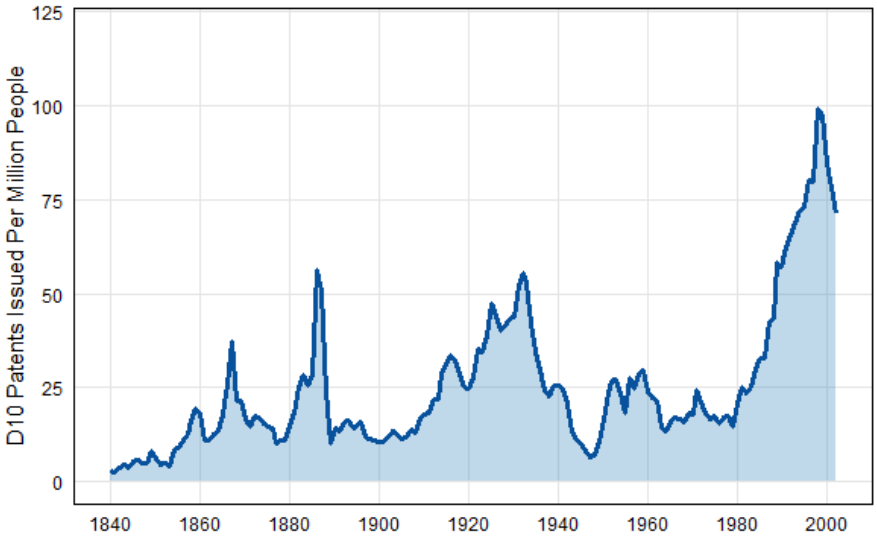
[51] Kelly et al, *supra* note 16. As the authors also discuss, there are several good reasons to believe these scores are a good measure of innovation (and the innovativeness of particular patents). First, the scores are highly correlated with citations, a traditional measure of the influence of a patent. But they have an important advantage—citation data is only available for recent years, while this measure goes back much further. Second, the scores match lists of historically important patents—these patents score highly on average. Third, the number of top-scoring patents issued by the Patent Office each year predicts productivity growth in the economy over both 5- and 10-year horizons. This is what we would expect if the scores are capturing important basic innovations—the patents issue, and then the inventions diffuse into the economy and boost productivity.

[52] Id.

[53] *Id.* at 316–18 & Appendix.

[54] *Id.*

Figure 8: Technology Waves, 1840–2002



Google PageRank, and (as noted above) Amazon's patent on one-click shopping.[55]

Strikingly, each of these major waves of basic innovation broadly coincides with one of the major waves of patent litigation—one in the late nineteenth century, another in the 1920s and 1930s, and a third in the 1990s and 2000s. This is a first clue that waves of innovation and litigation may be connected.

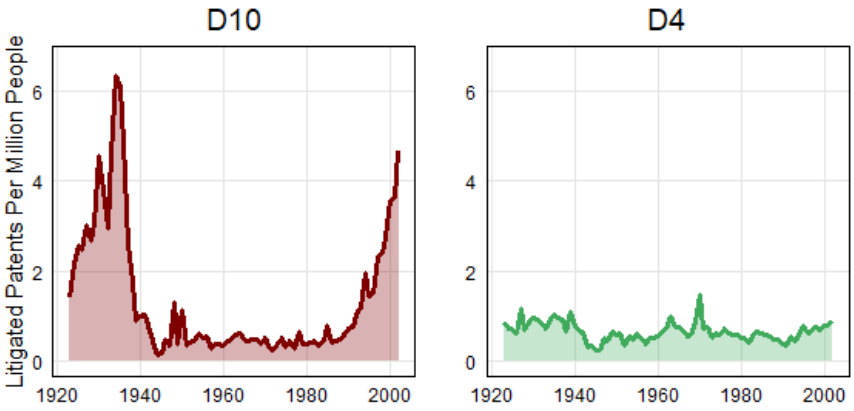
Next, the patent-level data in the litigation database shows that (at least during the period 1923–2002) waves of basic innovation and patent litigation do not merely coincide—top-scoring patents are highly litigated and explain a large share of the variation in the volume of patent litigation.

As an example to build intuition, Figure 9 on the next page begins by contrasting the litigation of patents in two particular deciles—D10 (the top-scoring decile) and D4 (a middle-scoring decile). In other words, the figure shows the number of patents from each decile (D10 and D4) that were asserted in litigation in each year from 1923–2002, scaled per capita.

The basic message of the results is easy to see in this figure. Litigation patterns in the two deciles looks quite different. The patent system has tended to experience a relatively steady stream of litigation involving low-scoring patents (D4), punctuated by

[55] *Id.*

Figure 9: D10 vs. D4, 1923–2002



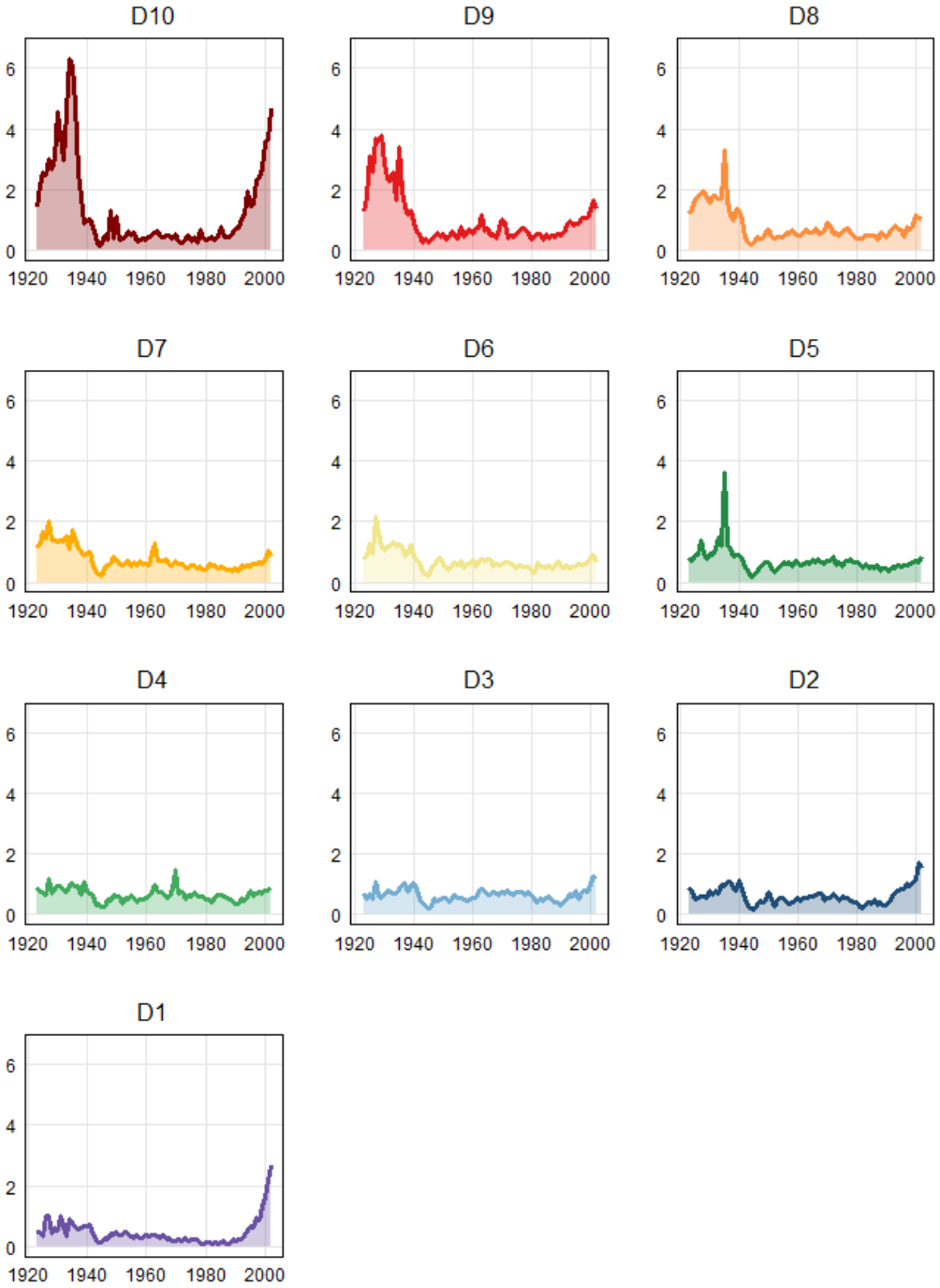
large surges of litigation over top-scoring patents (D10). And these surges drive litigation waves—the main thing that distinguishes high-litigation years from low-litigation years is the number of top-scoring patents asserted in litigation.

Figure 10 on the next page shows the analogous plots for all ten deciles, D1–D10. The same basic pattern appears—litigation in lower deciles is relatively flat, while litigation in higher deciles (particular D10) is highly volatile.

Of course, this is not to say that the lower deciles contribute nothing to the variance of patent litigation. For example, the year 1935 has an enormous spike of litigated D5 patents. This was due to an extensive campaign of patent litigation launched by a single patent owner, Samuel B. Smith, who asserted his egg-incubator patent against hundreds of different defendants.[56] Smith launched his campaign after the Supreme Court initially found his patent valid—only for the Court to later declare the patent invalid over newly-discovered prior art in a later appeal (perhaps the Court had second thoughts after seeing the results of its first decision).[57] There are other examples too. The data show a mini-wave of litigation of lower-scoring patents (D1) in the late 1990s and early 2000s. And all deciles contributed relatively equally to the dramatic fall in patent litigation during World War II. But, on the whole, top-scoring patents have played the larger role in driving changes in the volume of litigated patents.

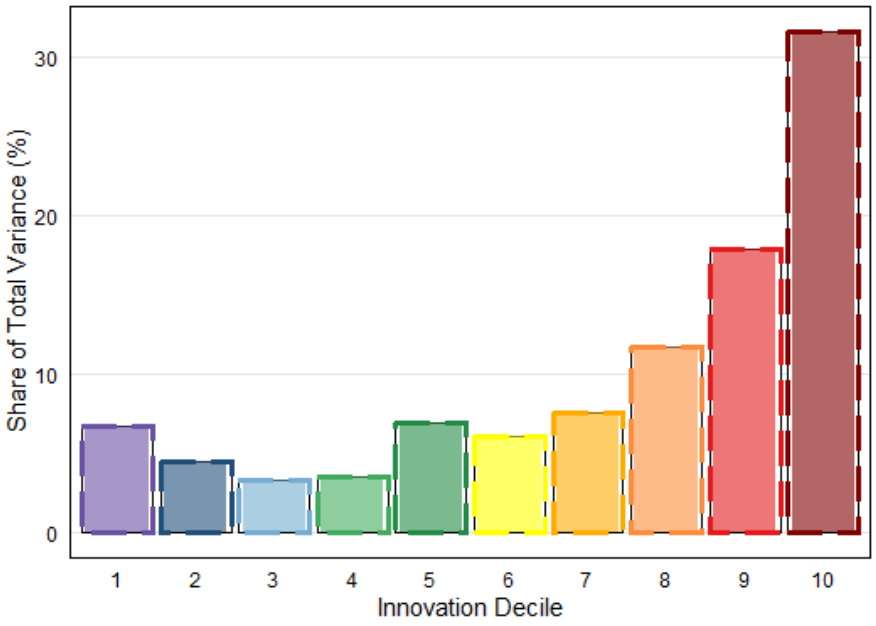
[56] *See* Katznelson, *supra* note 1.
[57] *See Smith v. Hall*, 301 U.S. 216 (1937).

Figure 10: Deciles 1–10, 1923–2002



A natural approach to quantify these relationships is a variance decomposition. This analysis decomposes the total variation in the volume of litigated patents into the contribution of each decile. The full details available in Appendix B. The results are shown below in Figure 10. During the period 1923–2002, top-decile patents by themselves explain more than 30% of the variance in litigation volume, and the top three deciles explain more than 60%. In short, top-scoring patents contribute most of the variance in litigation volume.

Figure 11: Variance Decomposition



B. Longevity of Top Patents

The data also provide insight into the feature of revolutionary patents that drives litigation. These patents explain a large share of litigation volume. But is there anything distinctive about their litigation dynamics?

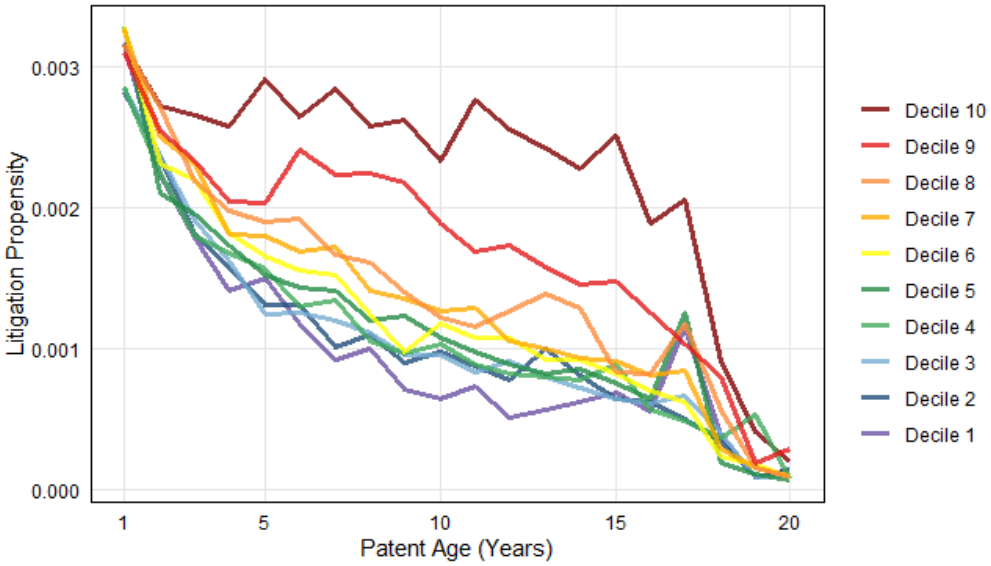
It is well-known that the age of a litigated patent is an important property. Specifically, Love (2013) shows that, on average, a patent is most likely to be litigated soon after it issues from the Patent Office, and then becomes less likely to be litigated over time.[58]

[58] *See* Brian Love, *An Empirical Study of Patent Litigation Timing*, 161 U. PENN. L. REV. 1309 (2013).

Accordingly, a good way to think about the question of what drives litigation is to examine the age/litigation profiles of patents in each decile. In other words, for patents in each decile, we can ask—how likely is the patent to be litigated in the first year after it issues from the Patent Office, in the second year, and so on—and then compare the results across deciles.

Figure 12 shows these age curves for each decile. The details of how the curves are constructed are available in Appendix B.

Figure 12: Age/Litigation Curves, 1923–2002

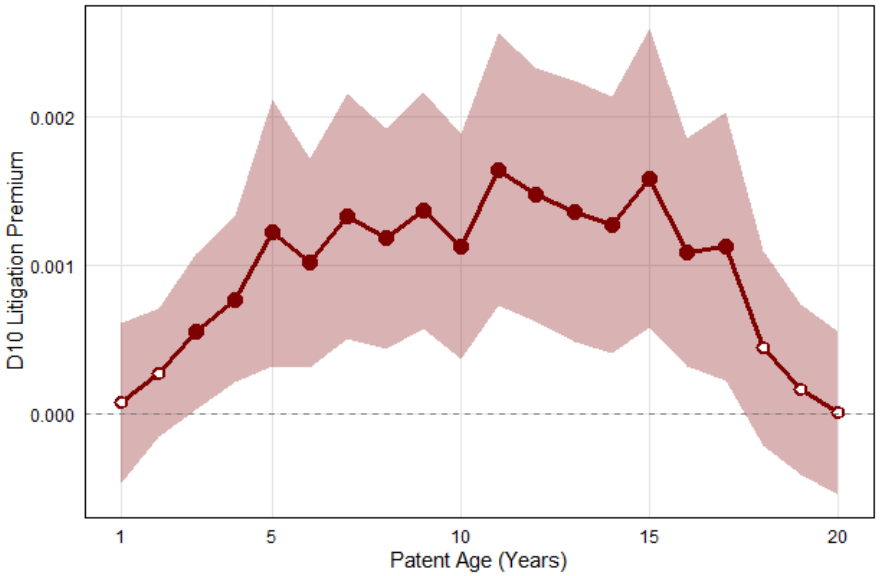


A surprising property jumps out from this figure. Top-scoring patents (particularly D10) are indeed litigated at greater rates—but this is driven almost entirely by longevity. At Age 1 (the first year after they issue from the Patent Office) patents in all deciles are litigated at relatively equal rates (~3 assertions per thousand patents). But, over time, the litigation propensities diverge. The litigation propensities of patents in lower deciles drop off relatively quickly, falling to around ~1 assertion per thousand patents after 10 years. But top-scoring patents (and particularly D10 patents) remain highly litigated for a longer time. Even fifteen years after they issue from the Patent Office, D10 patents are still highly litigated at ~2.5 assertions per thousand patents.

We can check whether these differences are statistically significant by testing whether membership in the D10 decile predicts higher litigation propensity at each age. The details of the regression are available in Appendix B. The results are shown

below in Figure 12, which plots the "D10 Premium" (the extra propensity for a D10 patent) at each age, alongside the confidence interval. Dots are bolded at ages where the litigation propensity is significant, and hollow at ages where it is not.

Figure 13: D10 Premium



The results match the qualitative observations from Figure 12. The D10 premium is low and non-significant in the early years (Ages 1–2), when all patents are litigated at similar rates. The D10 premium is also non-significant in the later years (Ages 18–20), after patent expiration (recall that a patent term generally runs for 17 years after the patent issues from the Patent Office). However, in the middle years (Ages 3–17), D10 patents look quite different—they have a large and significant litigation premium.

Examples of these dynamics appear in each litigation wave. For example, between 1997 and 2002, Acticon Technologies (a patent licensing company) filed more than a dozen lawsuits against major manufacturers of computer equipment (like IBM and Motorola). These lawsuits asserted several top-decile patents from the mid-1980s on a basic architecture for an internet modem. In other words, even 15 years later, this old design was still a live topic of patent litigation. Or, to take another example, in 2002 Agere (a semiconductor maker) filed a series of lawsuits against industry competitors asserting top-decile semiconductor patents that had been issued by the Patent Office in the early 1990s and late 1980s, between 9 and 17 years before the lawsuit was filed.

Similar dynamics appear in the early twentieth century litigation wave. For example, from 1934–1936 (the height of this wave), Radio Corporation of America (RCA)—a major player in the radio industry—filed more than two dozen lawsuits asserting U.S. Patent No. 1,297,188 against a wide variety of radio companies, including equipment manufacturers (like Amperite Corp.) as well as small retail operations (such as the Woodlawn Radio & Music Co.). This top-decile patent claimed a method of amplifying electrical currents, a central technological challenge in radio of the early twentieth century. It was also 16 years old at the time it was asserted in litigation—it issued from the Patent Office in 1919.

V. DISCUSSION

These results are stronger evidence that technological revolutions have played a large role in driving waves of patent litigation. Using the linked litigation database, we can look inside waves of patent litigation and see that they are substantially composed of top-scoring patents. The results also help to explain the origins of patent thickets.

An interesting observation is that, historically, complaints about patent thickets have seemed to focus on collections of patents surrounding revolutionary technologies. In the late nineteenth century, railroad operators were repeatedly sued on the "essential" components of their products, many of which had only recently been invented.[59] In the twentieth century, patent thickets grew in radio and synthetic chemicals, and legal commenters complained about the "hundreds of important patents to be reckoned with" in these "developing" technologies.[60] And the same dynamics have arisen once again in modern biotechnology, telecommunications, and computers.[61]

The data suggest a surprisingly simple story about what might be going on—why patent thickets form in particular decades, and why they seem to be connected to technological revolutions. The key is the behavior of top-scoring patents. Most patents issue from the Patent Office, are influential for a short time, and then fade as technology moves on. Accordingly, the stock of relevant patents turns over reasonably quickly. But high-scoring patents (and particularly patents in the top decile) work

---

[59] See Steven W. Usselman, *Patents Purloined: Railroads, Inventors, and the Diffusion of Innovation in 19th-Century America*, 32 Tech. & Culture 1047, 1063 (1991); Steven W. Usselman, REGULATING RAILROAD INNOVATION: BUSINESS, TECHNOLOGY, AND POLITICS IN AMERICA, 1840–1920, at 104–10 (2002); Beauchamp, *supra* note 1, at 929–30.

[60] *See* Petro, *supra* note 10, at 390.

[61] *See* Heller & Eisenberg, *supra* note 5; Shapiro, *supra* note 5.

differently. They issue, and then remain relevant for a longer time because they are embodied in future generations of technology.

The distinctive feature of a technological wave—such as the electrical innovations of the early 20th century, or the telecommunications revolution of today—is that many of these basic patents issue around the same time. And, as more and more basic patents accrete, new innovations become encumbered by "thickets" or "anti-commons" of long-lived rights. Companies that want to commercialize new products must worry about a long backlog of legal risks.

The Acticon cases illustrate the nature of the challenge. These lawsuits asserted top-decile patents—likely to cover a wide swath of future technology—that were issued in 1985. Yet, these patents were first asserted in litigation only in 1997 (and then in a flood of cases in the next 5 years). It is quite plausible that the products at issue in these cases had been designed without anyone at the defendant companies knowing about these patents. A complex network of overlapping patents on basic ideas creates landmines for future inventors.[62]

There are two other features of technological revolutions that make this problem especially thorny. One, when new categories of products are first created, almost every component has recently been invented, so there is very little in the public domain. In other words, any version of the product is likely to be covered by multiple patents. Railroads in the nineteenth century were sued on the "essential" components of their products, which had almost all been recently invented;[63] early radio sets needed to use "practically all of the inventions which were then known;"[64] and today many patentees have come forward claiming to have patents that cover the critical ingredients of e-commerce or video-on-demand.[65]

Two (and relatedly), technological revolutions are likely to be times when especially few *alternatives* exist, even patented ones. The problems created by a multi-component product are much less severe if each component has alternatives available—in other words, if a cell-phone maker can choose from multiple antenna designs, or a railroad has multiple options for the brakes. In that case, even if each alternative design is patented, no individual

---

[62] *See* Petro, *supra* note 10, at 390 ("It is often impossible to tell which patents are key patents in a complex and rapidly growing art").
[63] Beauchamp, *supra* note 1, at 929–30.
[64] Llanes & Trento, *supra* note 6.
[65] Mark A. Lemley, *Software Patents and the Return of Functional Claiming*, 2013 WIS. L. REV. 905, 907 (2013).

patent holder has excess leverage. If one patent-holder demands unreasonable terms, the manufacturer can switch to another design. However, technological revolutions are periods when the basics are discovered. These are ideas that every version of a product must use—which is why revolutionary patents continue to show up in litigation even 15 years after they issue. The absence of alternatives makes problems of fragmented rights especially challenging, by placing substantial leverage in the hands of each patent-owner.

Thinking about this problem from a policy perspective, there are two key points. First, not all top-scoring patents are created equal. Some reflect fundamental breakthroughs that were difficult to achieve, and likely could not have happened without the inducement of a patent. And if a patent "thicket" is the price of having the technology at all, then it is one worth paying. But this is not always true. Some top-scoring patents might capture an important idea that was not yet patented, but one that was nonetheless in the air, and likely to be discovered soon anyway. For example, a recent top-scoring patent from the internet revolution is Amazon's patent on one-click shopping.[66] That was surely an important (and influential) idea in e-commerce. But it is less clear that this was a case where the reward of a patent was essential, rather than a case where a number of natural ideas in e-commerce were in the air in the late 1990s and early 2000s.

Second, a well-known result about patent thickets is that, when they bite more severely, the patent system wants "weaker" rights—in other words, stricter standards that make patents more difficult to obtain and enforce.[67] In essence, when thickets are worse, each extra patent effectively imposes a larger externality on other inventors by contributing to bargaining frictions and multiple-marginalization problems (which make it harder for all inventors to get rewarded). Accordingly, the patent system wants to rely more on the chance that ordinary tinkering and non-patent incentives will be sufficient to discover the easier components of a multi-component product, and reserve patents for the hardest contributions.

Together, these points make out a case for time-varying patent law. As technological revolutions unfold—and waves of

---

[66] Kelly et al., *supra* note 16.

[67] *See* Shaprio, *supra* note 5; James Bessen & Eric Maskin, *Sequential Innovation, Patents, and Imitation*, 40 RAND J. ECON. 611 (2009); Gaston Llanes & Stefano Trento, *Patent Policy, Patent Pools, and the Accumulation of Claims of Sequential Innovation*, 50 ECON. THEORY 703 (2012); Gaston Llanes & Trento, *supra* note 6.

patent litigation rise and fall—there tend to be multi-decade periods where the problem of thickets is more or less severe. And, when thickets are worse, the patent system would like to cut back on the marginal patent.

As a matter of law, the patent system has a set of broad, core principles that are set forth in the patent statute and have remained relatively constant over time. However, these broad principles leave very substantial room to be applied more strictly or more leniently. For example, one basic rule is that a patentable invention must be sufficiently creative ("nonobvious"). But how much creativity is enough? And what kind of evidence is relevant? There are many cases where the evidence is somewhat ambiguous—there are hints of the invention in the prior art, and a sense that the idea might have been in the air, but nothing concrete—so a decision-maker might go either way. And there are many other examples of this character, e.g., whether the scope of an ambiguous patent is construed broadly or narrowly, or whether strong remedies (injunctions) are readily available in patent suits.

It is difficult for the Patent Office to act on this insight—given limited resources, it has little choice but to err in favor of granting patents in debatable cases. But there is a role for courts, who oversee patent enforcement. Indeed, enforcement is where the risks of patent thickets ultimately arise. Litigation is the backdrop for licensing negotiations, and a surprise lawsuit is the penalty for a missed patent. If judges treat litigated patents more strictly—by, on the margin, ruling against patents of doubtful validity, or construing ambiguous patents narrowly—that can help to clear the thickets by reducing litigation risk, and allocating the rewards of the patent system to the inventors who have solved the hardest problems.

Intriguingly, courts have always done some version of this (if roughly and imperfectly). In fast periods, they have tended to look more strictly on litigated patents. For example, in the late nineteenth century, the Supreme Court gave a narrow construction to patents on railroad air brakes (including Westinghouse's famous patents), worrying about the future path of progress in the field.[68] In the twentieth century, the Supreme Court (over a bitter dissent) found Marconi's wireless radio patents invalid over the prior art.[69] And, recently, the Supreme

---

[68] *See Railway Co. v. Sayles*, 97 U.S. 554 (1878); *Westinghouse v. Boyden Power Brake Co.*, 170 U.S. 537 (1898).

[69] *Compare Marconi Wireless Co. v. United States*, 320 U.S. 1, 10 (1943) (majority) ("Long before Marconi's application for this patent the scientific principles of

Court has cracked down on patenting in computers and biotechnology, worrying that a loose approach might stifle creativity, rather than enhance it.[70] These are not isolated examples, but rather parts of broader trends—what patent-law commenters have often called the "swinging pendulum" of judicial attitudes toward patent rights.[71]

This behavior by courts has sometimes come in for heavy criticism. After all, patent law is a field of property. It is natural to think that the role of courts should be much simpler—to keep the law "stable and clear."[72] Indeed, for nearly the entire history of the patent system, there have been calls for the law to achieve greater stability. For example, the leading patent-law treatise of the late nineteenth century argued that "not merely the importance, but the absolute necessity" of a stable theory of patent rights was "demonstrated by the remarkable changes in the attitudes of our courts toward patentees during the past few years," swinging from "an extreme liberality" to an "equally extreme strictness."[73] And similar calls have continued unabated for more than a century.[74]

Yet, stability in patent law has remained elusive. Each generation of judges has felt compelled to make changes to the way they assess litigated patents, under the pressures of litigation surges or busts. Some courts have steered toward a strict approach, worrying that leniency in patent law "tends rather to obstruct than to stimulate invention" and "lay[s] a heavy tax upon the industry of the country."[75] Others have found those worries difficult to understand, and focused on making sure that inventors (a "meritorious class") receive the benefit of the doubt when they arrive in court.[76]

---

which he made use were well understood and the particular appliances constituting elements in the apparatus combination which he claimed were well known") *with id.* at 62 (Frankfurter, J., dissenting) ("The inescapable fact is that Marconi in his basic patent hit upon something that had eluded the best brains of the time working on the problem of wireless communication").

[70] *E.g., Bilski v. Kappos*, 571 U.S. 593 (2010); *KSR Intern Co. v. Teleflex Inc.*, 550 U.S. 398 (2007); *Mayo Collab. Serv. V. Prometheus Labs., Inc.*, 566 U.S. 66 (2012).

[71] Mark A. Lemley, *The Surprising Resilience of the Patent System*, 95 TEX. L. REV. 1, 13–14 (2016); ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS 79 (2007).

[72] *Bilski*, 561 U.S. at 613 (Stevens, J. concurring in the judgment) ("In the area of patents, it is especially important that the law remain stable and clear.").

[73] 1 WILLIAM C. ROBINSON, THE LAW OF PATENTS FOR USEFUL INVENTIONS 36, § 23 n.1 (1890).

[74] *See, e.g.,* Lee Petherbridge, *Patent Law Uniformity*, 22 HARV. J. L. & TECH. 421 (Spring 2009) (colleting arguments).

[75] *Cuno Engineering Corp. v. Automatic Devices Corp.*, 314 U.S. 84, 92 (1941).

[76] *See Commissioner of Patents v. Whitely*, 71 U.S. 522, 532 (1867); *Union Carbide Corp. v. American Can Co.*, 724 F.2d 1567, 1574 n. 4 (Fed. Cir. 1984).

Judges have largely been reacting to the circumstances in front of them, and have not tried to offer a broader theory that explains what is going on. But we can see a logic behind these choices. Patent law faces different kinds of property rights problems over time, so there is a role for the law to dynamically adapt the scrutiny given to litigated patents.

VI. CONCLUSION

The patent system has experienced several large waves of litigation, but their origins and implications remain incompletely understood. To study them, the paper uses a large language model to digitize decades of Patent Office records, and links them to a quantitative measure of innovation.

The results point to technology as an important factor. Particularly, they relate to a classic idea in the history of technology. Economic historians have often argued that technological progress does not occur at a smooth, constant rate over time, but rather is highly variable.[77] In some decades, major breakthroughs occur—"macroinventions" or "general purpose technologies"—that open the door for additional innovation, and progress moves rapidly. In other decades, technology is more stagnant.

The results of the paper suggest that this basic fact about technological progress has important implications for patent law. Technological revolutions play a large role in explaining waves of patent litigation. They contribute to the formation of patent thickets. And they imply that the optimal approach of the law plausibly varies over time as well.

---

[77] *See, e.g.*, Joel Mokyr, *Punctuated Equilibria and Technological Progress*, 80 AM. ECON. REV. 350, 350–52 (1990); Timothy F. Bresnahan & M. Trajtenberg, *General Purpose Technologies 'Engines of Growth'?*, 65 J. ECONOMETRICS 83, 83–84 (1995).

APPENDIX A: DATABASE

This Appendix covers the litigation database. Part A reproduces the prompt. Part B describes the data cleaning.

A. Prompt

Hi! I would like your help with a data entry project. I have uploaded a PDF document (below) with pages from the PTO Gazette. It contains a list of patent suits (among other things). And each entry on the list may contain the following pieces of information about the patent suit: (1) docket number, (2) filing date, (3) decision date, (4) notice date, (5) plaintiff, (6) defendant, (7) court, (8) asserted patents.

I would like you to read the document, find the list of patent suits, find the relevant pieces of information for each suit in the list, and create an artifact that contains the information in the form of a pipe-delimited markdown (not raw) table with rows and columns. I would like one row for each suit, and one column for each piece of information, except I want separate columns for up to 20 possible patents associated with each suit (e.g., "Patent 1", "Patent 2", "Patent 3", ... "Patent 20"). I'd like the columns in the order listed above—docket number, then filing date, etc. Please include "Doc." at the start of each docket number.

A few notes:

Every single page of the PDF has patent suits (if there are multiple pages, the list runs across all pages). Make sure to catch every page.

Some list entries might lack a docket number for the suit. That's okay — they are still suits, and you should still include a row for them (just leave the docket number field blank). Follow the same approach if a suit is missing any other piece of information; just leave that field blank.

Some suits may involve many patents—possibly 10, 15, or even 20 separate patents. Please be sure to grab all patents associated with each suit. Also, please include all kinds of patents, including reissue patents (Re.), design patents (Des. or D.), and plant patents. If there are trademarks (T.M. or Reg. No.), grab those too.

Some entries may contain multiple docket numbers. That means multiple suits were filed involving the same patents (each docket number is a separate suit). Please give each docket number its own row. Also, in these entries, the gazette may indicate that some details for later docket numbers are the same as for an earlier docket number—for example, “decree as above” or “same.”

Some entries are nothing more than a patent number followed by a cross-reference to another patent number in parentheses. For example: “954,290. (See 902,584.).” You can ignore those. They are not independent suits, but rather just cross-references to other suits.

Please enter dates in the format YYYY-MM-DD.

Please be accurate about the dates — correctly classify whether a date is a filing date, decision date, or notice date. Here’s how to distinguish them:

Filing Date: The date the suit (or appeal) was filed in court. These dates will explicitly mention filing. Examples: "filed Dec. 1, 1927” or “suit filed April 19, 1927” or “appeal filed Oct. 25, 1929.”

Decision Date: The date a court decision was made or the case was otherwise resolved, such as a decree, dismissal, affirmation, judgment, or discontinuance. Examples: “Decree for injunction May 18, 1927" or "Dismissed June 1, 1928" or “Affirmed Feb. 6, 1930” or “Discontinued October 28, 1931” or “Decree finding infringement of claims 1, 2, and 3 and sustaining patent, March 9, 1929.”

Notice Date: The date the Patent Office received notice of a court decision (rather than the date of the decision itself). These have a description of the decision followed by the notice date enclosed in parentheses, and will always be explicitly labeled as a “notice.” Examples: “Dismissed (notice June 1, 1928)” or "Decree sustaining patents and granting injunction (notice dated Mar. 21, 1928)” or “Order of discontinuance (notice Dec. 10, 1929).”

If a date (or any other detail) is ambiguous or unclear, please indicate this in the output so I can review it manually. Please be careful. If you need to think more to get something right, that’s okay — my #1 goal is accuracy. The document is attached.

B. Data Cleaning

Fields that should have a standard format—docket number, dates, court, and patents—were standardized.[78] Several categories of entries were also dropped: (1) entries without a filing date, (2) entries for suits in appellate court (rather than district court), and (3) entries without any patents (including trademark cases, which are occasionally reported in the Gazettes).

The more challenging issue was de-duplication. The Gazettes contain a fairly substantial number of duplicate entries (multiple reports for the same case) simply because of the structure of the reporting obligation. The patent statute requires clerks of court to file reports when a case is filed or decided (including pleading amendments).[79] Accordingly, many events in a case can potentially generate a report—a case is filed, the pleadings are amended, a preliminary injunction issues, final judgment is entered, the mandate is returned after appeal, and so on.

Identifying and dropping these duplicate reports is essential to correctly counting the number of suits. And a careful, rigorous de-duplication process is especially important due to a set of changes in the Gazette reporting practices over time.

The early Gazettes (1920s to early 1950s) generally report just one date per entry (either a filing date or a decision date). Accordingly, the data-cleaning step of dropping entries without a filing date will tend to eliminate a very substantial share of the duplicates. However, in the early 1950s, the Gazettes began including a filing date with *every* entry—even for notices that were sent later for other events. And, for its part, the LitAlert data contains a single field that is used to record *all* dates (filing or decision). Accordingly, dropping entries without a filing date will eliminate very few duplicates after the early 1950s.

The unfortunate result is that the data from later years will tend to have many more duplicates, so a weak de-duplication process will tend to introduce variation in the number of duplicates over time.

A careful de-duplication effort needed to overcome two challenges. One is volume. The original Gazette dataset—even after dropping cases without a filing date—has ~53,000 entries,

---

[78] There is a limit to the standardization of docket numbers. It is possible to fix easy issues, like making sure blank fields have a consistent format, or handling the small number of rows that have multiple docket numbers for various reasons. But there is a huge volume of formatting variation that defies easy resolution. For example, the same docket number might appear as "70-cv-123" vs. "70-123" vs. "70/123" vs. "70-cv-00123." These issues were handled during de-duplication.

[79] R.S. § 4921 (1922).

and the LitAlert dataset itself has ~29,000 entries. If we imagine every pair as a possible set of duplicates, that is more than a billion total comparisons.

The other challenge is subtle variation in the data. In general, there are six useful fields of information for each case: (1) docket number, (2) filing date, (3) plaintiff, (4) defendant, (5) court, and (6) patents. It is easy to drop exact duplicates. But many pairs of duplicates may differ on some fields due to, e.g., spelling errors, differences in the formatting of docket numbers, or different abbreviations in party names.

Two basic steps were used. First, exact duplicates were dropped. For the Gazettes, that meant exact duplicates on all six fields. For LitAlert, this needed to be relaxed, since filing date is not a useful field. Instead, it meant exact duplicates on the other five fields. Another feature of the LitAlert data provides extra reassurance. In the early 1970s, federal courts switched to modern docket numbers, which are both longer and include the filing year of the case, so they are extra likely to be unique. For example, a case filed in 1930 might have a docket number like "E 400," while modern cases have docket numbers that look like "4:99-CV0148-HLM." It is not realistically possible for two modern cases to be an exact match on docket number, plaintiff, defendant, court, and patents without being duplicates. In total, ~5,000 exact duplicates were dropped from the Gazettes, and ~4,300 exact duplicates were dropped from LitAlert.

Second, candidate "fuzzy" duplicates were identified for manual review, i.e., pairs that are not an exact match along at least one dimension. To do this, matching scores from 0 to 1 were created for each of the six available fields:

1. Docket Number. A score was created using Jaro-Winkler similarity, which measures the distance between two strings of characters. Scores run from 0 to 1 (1 is a perfect match). A score of .7 or less is an extremely weak signal, so scores were given used a scaled metric:

$$\begin{cases} 0 & \text{if } JW_{sim} \leq .7 \\ \frac{(JW_{sim} - .7)}{.3} & \text{otherwise} \end{cases}$$

2. Filing Date. 1 if a match, 0 otherwise.

3. Plaintiff. Standardized versions of plaintiff names were created, e.g., harmonizing treatment of abbreviations like "Inc." and "Incorporated." Scores were then given using the scaled JW metric.

4. Defendant. Similarly, standardized versions of defendant names were created, and scores were given using the scaled JW metric.

5. Court. 1 if a match, 0 otherwise.

6. Patents. Scores were given using Jaccard similarity for the two lists of patents asserted in each suit (intersection over union).

Summing the scores gives a total score from 0 to 6 for each pair. High-scoring pairs are more likely to be duplicates, and therefore good candidates for manual review.

Ideally, we would compute scores for all pairs. However, with more than a billion total comparisons, this is computationally infeasible. Some way is needed to identify smaller blocks of useful comparisons to make.

Three main blocking strategies were used: (1) by filing date, (2) by lead patent, and (3) by court. For example, if blocking by filing date, we examine only pairs with the same filing date, and calculate scores for them. For LitAlert, filing date was not a very useful field, so as extra insurance a run was also done blocking by docket number (which proved not to be necessary).

Blocking dramatically cuts down on computation, but is also very likely to pick up all strong candidates. The blocking strategies use sensitive details that have a unique correct format, which are particularly likely to be recorded correctly. And multiple blocking approaches are used to catch errors. For a duplicate pair to be missed, errors would need to arise in three separate critical fields.

Once a set of scores was calculated, manual review was done of all pairs that were promising candidates. In particular, pairs were reviewed that had either a high overall score (5+ Gazettes, 4+ LitAlert) or a medium overall score (3.3+ Gazttes, 3.0+ LitAlert) plus at least one other promising feature: (1) defendant score above 0, (2) plaintiff score below 1, (3) docket score above .9, (4) court score below 1, or (5) patent score below 1. Finally, for the Gazettes, candidate pairs from the same issue were dropped (as single issues do not print duplicates).

Two notes. First, lower thresholds were used for LitAlert as those scores were effectively out of five rather than six. Bottom thresholds were set by experience, where the flow of duplicates dried out. Second, the list of promising features might seem a little odd (why is it good to have a *low* plaintiff score?), but there is a unifying explanation for why they are good signals, discovered through experience.

The key point is that—given litigation filing patterns—it really matters *where* points are lost. Two suits that have an only-medium score because they got zero on the docket number and defendant (indicating large difference in these fields) are overwhelmingly unlikely to be duplicates. That is because it is very common for a plaintiff to file multiple lawsuits against different defendants, asserting the same patent. By contrast, two suits that lost points because they have different *plaintiffs* are much more promising—it is very unusual for a single defendant to be sued on the same patent by different plaintiffs. In that case, there is likely an error or other inconsistency in the plaintiff fields. A similar logic holds for the patent and court fields. In sum, the set of special conditions finds pairs of suits that lost points in promising places.

In total, manual review was done of ~16,000 pairs of potential fuzzy duplicates. ~6,000 fuzzy duplicates were dropped from the Gazettes, and ~2,000 fuzzy duplicates were dropped from LitAlert.

APPENDIX B: ANALYSIS

This Appendix reports the details of the analysis. Part A discusses the variance decomposition, and Part B discusses the age/litigation curves.

A. Variance Decomposition

The variance of the total number of patents asserted in litigation each year, per million people, over the period 1923–2002 is the sum of the variances and covariances of the ten deciles:

$$Var(L) = \sum_{d} Var(L_d) + \sum_{d \neq d'} Cov(L_d, L_{d'})$$

The full variance/covariance matrix is:

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D1	.200	.106	.063	.039	.046	.052	.066	.104	.161	.439
D2	.106	.084	.050	.035	.044	.035	.046	.075	.096	.279
D3	.063	.050	.049	.037	.041	.039	.044	.061	.082	.159
D4	.039	.035	.037	.050	.050	.046	.055	.079	.114	.173
D5	.046	.044	.041	.050	.163	.079	.106	.181	.225	.394
D6	.052	.035	.039	.046	.077	.102	.106	.146	.238	.318
D7	.066	.046	.044	.055	.106	.106	.143	.184	.292	.407
D8	.104	.075	.061	.079	.181	.146	.184	.298	.434	.662
D9	.161	.096	.082	.114	.225	.238	.292	.434	.757	1.012
D10	.439	.279	.159	.173	.394	.318	.407	.662	1.012	2.183

B. Age/Litigation Curves

There are 16,000 propensity data points in total: 10 deciles x 20 ages x 80 years of litigation. Each data point is the propensity of a particular cohort of patents (issue year and decile) observed in a particular year of litigation. For example, in 1924, we observe the cohort of D10 patents issued in 1923 being litigated at Age 1. To obtain the propensity, the total number of assertions is divided by the size of the cohort.

Two notes. First, terminology. Available data shows the *calendar* year in which patents were issued, but the *fiscal* year in which they were litigated. Age 1 is defined as first full fiscal year after a patent is issued. For example, if a patent issued in calendar year 1923, then fiscal year 1924 is Age 1.

Second, cohorts. Litigation data is contributed by patent cohorts issuing between 1903 and 2001. For example, for patents issuing in 1903, we observe the Age 20 propensity in the first year of the litigation data, fiscal year 1923. Likewise, for patents issuing in 2001, we observe the Age 1 propensity in the last year of the litigation data, fiscal year 2002.

From this data, age/litigation curves are straightforward to calculate. There are 80 data points for each age-decile pair, and averages are taken.

Finally, to test the significance of the D10 premium, we regress litigation propensity on the combination of age and decile:

$$propensity_i = age_i \times D10_i + \varepsilon_i$$

D10 is a dummy variable that is equal to 1 if $patent_i$ is in D10, and 0 otherwise. Accordingly, the results show two age curves—the baseline age curve for deciles D1–D9, and a premium for D10 at each age.

A potential concern is time correlation. The data for each age is effectively a time series—we observe litigation propensities in 1923, 1924, and so on. It is plausible that these propensities serially correlated over time. For example, high litigation in 1930 is followed by high litigation in 1931, possibly driven by common factors. Accordingly, autocorrelation (and heteroskedasticity) robust standard errors were calculated.

Results are shown in Table 4 below. There are two basic messages.

The first column of Table 4 shows the litigation curve for deciles D1–D9. The Age 1 coefficient shows the baseline litigation propensity at Age 1, and Ages 2–20 show the change over Age 1 in each year. As expected, Age 1 is positive (patents are likely to be litigated in the first year), and Ages 2–20 are negative (patents become less likely to be litigated as they age).

The second column in Table 4 shows the D10 premium at each age. D10 patents are not significantly more likely to be litigated in early years (Ages 1–2) or later years (Ages 18–20), but they are significantly more likely to be litigated at Ages 3–17.

Table 2: D10 Premium

	D1–D9 Propensity	D10 Premium
Age 1	.00308***	.00007
	(.00008)	(.00027)
Age 2	-.00071***	.00027
	(.00007)	(.00022)
Age 3	-.00105***	.00055*
	(.00008)	(.00027)
Age 4	-.00134***	.00077**
	(.00008)	(.00028)
Age 5	-.00147***	.00122**
	(.00008)	(.00046)
Age 6	-.00152***	.00102**
	(.00009)	(.00036)
Age 7	-.00163***	.00132**
	(.00009)	(.00042)
Age 8	-.00176***	.00118**
	(.00010)	(.00038)
Age 9	-.00190***	.00137***
	(.00010)	(.00041)
Age 10	-.00195***	.00113**
	(.00010)	(.00039)
Age 11	-.00203***	.00164***
	(.00009)	(.00047)
Age 12	-.00208***	.00148***
	(.00009)	(.00044)
Age 13	-.00210***	.00136**
	(.00010)	(.00045)
Age 14	-.00215***	.00127**
	(.00009)	(.00044)

Age 15	-.00223***	.00159**
	(.00009)	(0.00051)
Age 16	-.00235***	.00109**
	(.00009)	(.00039)
Age 17	-.00222***	.00113*
	(.00014)	(.00046)
Age 18	-.00269***	.00044
	(.00010)	(.00033)
Age 19	-.00291***	.00016
	(.00009)	(.00029)
Age 20	-.00296***	.00001
	(.00008)	(.00028)