Dynamics of Patent Precedent and Enforcement: An Introduction to The UGA Patent Litigation Datafile

Matthew D. Henry
Department of Economics, Cleveland State University

Thomas P. McGahee
Analysis Group, Inc.

John L. Turner†
Department of Economics, University of Georgia

March 2013

Abstract
This paper describes a data set built to analyze patent litigation in the United States during 1929-2006. The data include all patents in published district court and appellate court decisions in which validity and/or infringement are at issue. We collect numerous patent-specific variables and litigation-specific variables for these patents. The data also include a set of random patents matched to the litigated patents by issue date and technology. We discuss the characteristics of the data, present some descriptive statistics, and describe research using the data.

JEL Code: K2, O3
Keywords: data, infringement, litigation, patent, validity

†Corresponding Author. Department of Economics, University of Georgia. Brooks Hall 5th Floor, Athens, GA 30602-6254. Phone: 706-542-3682. E-mail: jlturner@uga.edu. This project was generously supported by National Science Foundation grant SES-0751661.
1. Introduction

Recently in the United States, rates of patenting and of patent litigation have surged. These trends, particularly the acceleration in (costly) litigation, have led economists (e.g. Jaffe and Lerner 2004) and legal scholars (e.g. Lemley, Lichtman and Sampat 2005; Bessen and Meurer 2008) to argue that the US patent system is failing to achieve its Constitutional mandate “to promote the progress of science and useful arts,” and to propose policy changes. Their main argument is that too many patents are issued with too strong a presumption of validity, yielding too-frequent disputes that impose unnecessary costs on innovation.

The court system is clearly a key institution for influencing patent enforcement, through changes to either policy or precedent, thereby affecting the costs and benefits of innovation. For example, when the Court of Appeals for the Federal Circuit (CAFC) was established as the sole appellate court in United States patent cases in 1982, it adopted the “clear and convincing evidence” standard for patent validity instead of the weaker “preponderance of the evidence” standard that prevailed in most appellate courts prior to 1982. This led to an immediate, sharp reduction in the aggregate rate of invalidity in the United States in 1982 (Henry and Turner 2006), which itself led to the recent surge in patenting (Hall 2005).

The clear connection between the CAFC’s changes to court precedent and subsequent rates of patenting stand as a relatively rare instance of empirical research identifying the timing and effect of changes to policy and court precedent. Given contemporary controversies over patent enforcement, it is unfortunate that more work has not been done. Policy proposals geared towards changing the rate of patent enforcement are ideally interpreted in light of statistical evidence on patent litigation outcomes associated with different statutory approaches to enforcement. Yet, while numerous changes to patent law since the 1930s have affected patentability, damages, antitrust implications of patents, judicial procedure, and other features, few scholars have attempted to identify the impact of these changes on empirical rates of enforcement. The major reason for this is the lack of a comprehensive data set of US patent litigation. Alternative data sets (e.g. Janicke and Ren 2006; Moore 2001; Koenig 1980; Federico 1956) span relatively short periods and are often gathered specifically to avoid including years with key policy changes. The authors also

use different sources and different criteria for variables, so the sets are not easily combined.

The UGA Patent Litigation Datafile addresses these shortcomings with a set of virtually all patents found “invalid,” “not infringed” or “valid and infringed” by a district court (or the International Trade Commission) in a case where at least one decision (district court or appellate court) is recorded in the *United States Patents Quarterly* (USPQ) and where validity and/or infringement is a decisive issue.\(^2\) To these cases we match patent-specific variables. The two major sources of information are primary: complete opinions from patent decisions published in the USPQ, and the patent documents themselves. The USPQ is an ideal source for detailed, longitudinal data on patent litigation. It includes the entire opinions for virtually every appellate court decision and for a large sample of district court decisions. It has also maintained its focus on intellectual property cases and has kept the same publisher since 1929.

The longitudinal nature of the data set permits careful identification of key empirical stylized facts about the impact of policy changes, as well as tests of theories relevant to policy. These data have already yielded three publications (Henry and Turner 2006; Atkinson, Marco and Turner 2009; Henry forthcoming) and several working papers (McGahee 2011; McGahee and Turner 2013; Henry and Turner 2013) that shed considerable light on patent enforcement, firm behavior in litigation, patent value, and knowledge spillovers.

In considering intertemporal trends in such characteristics, these data are potentially useful for assessing the impact on patent enforcement of numerous specific changes to patent law, policy and legal precedent since 1929. These include but are not limited to (1) the vigorous US antitrust policy stemming from the hearings and reports of the Temporary National Economic Committee (TNEC) in the late 1930s, (2) the merging of the courts of law and equity in 1938, (3) the “Flash of Genius” test for patent validity promulgated by the Supreme Court in *Cuno Engineering Co. v. Automatic Devices Co.* [314 US 84 (1941)], (4) the Patent Act of 1946, which establishes “reasonable royalty” damages as a minimum level of damages and removes “unjust enrichment” as a means to calculate damages, (5) the Supreme Court’s establishing of the modern-day doctrine of equivalents in *Graver Tank & Manufacturing Co. v. Linde Air Products Co.* [339 US 605 (1950)], (6) the Patent Act of 1952, which rules out the “Flash of Genius” test (item 3) and establishes the statutory language for the modern-day presumption of patent validity, (7) the Supreme Court’s establishing of the modern-day interpretation of patent non-obviousness in *Graham v. John Deere & Co.* [383 US

\(^2\)http://www.terry.uga.edu/jltur/ntlpatentdata/.
1 (1966)], (8) the precedent that a finding of patent invalidity in one case creates a collateral estoppel from relitigating validity in subsequent cases, by the Supreme Court in *Blonder-Tongue v. University of Illinois Foundation* [402 US 313 (1971)], (9) the establishment of the patentability of living microorganisms and computer software in, respectively, the Supreme Court decisions in *Diamond v. Chakrabarty* [447 US 303 (1980)] and *Diamond v. Diehr* [450 US 175 (1981)], and (10) the establishment of the CAFC and its innovations to patent law. This paper discusses some of these changes in greater detail, but the data are useful for analyzing the impact of all of them.

The data also include a matching set of random patents. Building and improving upon Lanjouw and Schankerman (2001), these data will permit the study of differences in litigated and random patents, and of selection into patent litigation. In the context of the full 1929-2006 set, this addition will allow researchers to study the impact of policy changes on which patents are litigated. In work just underway, Henkel, Turner and Zischka (2013) are using these data to predict the share of patents that, if litigated, would be found invalid.

2. Literature Review

During the past fifty years, researchers have performed a sizable number of notable empirical studies of patent litigation.\(^3\) Two of the more comprehensive studies are primarily a written statistical record. At the request of a US Congressional subcommittee, Federico (1956) reports data on 1925-54 patent decisions published in the *United States Patents Quarterly* (USPQ). Koenig (1980), also using the USPQ, records slightly different statistics on patent decisions from 1953-78. These studies reach few substantive conclusions.

Other studies draw more conclusions from their data. Studying data from the *Official Gazette of the Patent Office* (1921-1960) and the USPQ (1961-73), Baum (1974) argues patent validity rates dropped between the early 1930’s and 40’s, but he estimates neither the year nor the significance of this drop. Dunner (1985) and Dunner, Jakes and Karceski. (1995) study decisions by the CAFC during 1982-94 and show that the CAFC affirms judgments in favor of patent owners more often than it affirms judgments in favor of accused infringers. They also find that the court was more likely to uphold findings of “not infringed” than it was to reverse them.\(^4\) Allison and Lemley (1998) examine validity rulings during 1989-96 and perform a battery of hypothesis tests on the

\(^3\)See Henry and Turner (2006) for more details about related literature.

\(^4\)See also Coolley (1989).
data as one group. They find that 54% of patents adjudicated were decided “valid.” Moore (2000; 2001; 2003) analyzes data gathered from the Administrative Offices of the US Courts on patent litigation since 1983. She estimates differences in outcomes of judge vs. jury trials (2000), the incidence of forum shopping (2001) and differences in outcomes for domestic vs. foreign patentees (2003). Her data do not include information about validity or infringement. Marco (2005) also uses USPQ data from 1977-97 to estimate probabilities of erroneous findings of validity and invalidity by district and appellate courts. Ball and Kesan (2006) study the final disposition for all cases filed in 1995, 1997 and 2000, and find that patents are rarely invalidated. Janicke and Ren (2006) study all cases decided by the CAFC during 2002-04 and find that patentees win roughly one in four cases. Schwartz (2011) shows that the CAFC received more appeals involving claim construction after Markman v. Westview Instruments, Inc. (52 F.3d 967 [Fed. Cir. 1995]) (1997-2003) than before (1991-94), that it drafted a greater percentage of written opinions, and that it analyzed the doctrine of equivalents less often.

Lanjouw and Schankerman (2001) study patents litigated between 1979-95. Their primary data, from Derwent Litalert, include cases not litigated to final decisions and cover between 50-70% of all cases during most years. Lanjouw and Schankerman (2001) capture a matched set of random patents and study differences in characteristics of litigated versus matched patents. They find that litigated patents have significantly more claims than random patents, are less likely to be of foreign origin, receive more forward citations and are generally more valuable.5 The Derwent data include a broader cross-section of cases than our data, but they do not include decisions. Additionally, they include patent numbers for some, but not all, of the patents in the cases. Derwent data continue to be collected and can be found via Westlaw.

Using entirely different data, Allison, Lemley, Moore and Trunkey (2004) study characteristics of litigated and non-litigated patents. Their data combine a comprehensive set of patents litigated in cases concluding in 1999-2000 with all patents from the original NBER patent datafile of Hall, Jaffe and Trajtenberg (2001).6 In general, they confirm the key findings of Lanjouw and Schankerman (2001) for more recent data, but they notably consider several additional variables (such as non-prior-art citations and post-issue assignment of the patent) and measure technology field differently. One of the authors, Kimberly Moore, identified the litigated patents by collecting

---

5See also Lanjouw and Schankerman (2004).
6The authors also select a smaller group of patents issued during 1996-98 for more intensive study.
all patent cases for 1999-2000 from the Administrative Office of the US Courts, then searching related documents (e.g. docket sheets, complaints, etc.). A comprehensive set of Administrative Office data can be found in the Federal Judicial Center’s “Federal Court Cases Integrated Database,” archived by the Inter-University Consortium for Political and Social Research at the University of Michigan (#8429). The main data span 1970-2000. Like Derwent data, these data include one particularly useful variable that USPQ opinions lack, filing dates.

Two additional archives retain detailed patent litigation records. Mark Lemley and Joshua Walker led (in collaboration with the Stanford Computer Science Department) the creation of The Intellectual Property Litigation Clearinghouse (IPLC) at Stanford University. The IPLC covers all lawsuits filed in US district courts since January 1, 2000 for patent infringement, manifest copyright, manifest trademark and manifest antitrust, and certain trade secret lawsuits. In addition, comprehensive records of recent patent litigation filings are available (for fees) through the Public Access to Court-Electronic Records (PACER) archive.

Because no existing data sets had the characteristics for comparing courts’ treatment of patents across the pre-CAFC and CAFC eras, Henry and Turner (2006) began the collection that became the UGA Patent Litigation Datafile. The goal was to capture large numbers of decisions from both before and after 1982, then perform time series analysis of frequencies of types of decisions, rates of appeal, and frequencies of appellate court decisions. Previous data sets cannot be combined because the data come from different sources and the authors use different criteria. In addition, many of these data sets were not constructed for the purposes of statistical analysis.

Since first putting the data together, Henry and Turner (2006; 2013) have used them to study rates of enforcement, first over 1953-2002 and then over 1929-2006. Atkinson, Marco and Turner (2009) use the data to show that there was significant non-uniformity in validity outcomes across circuits prior to the CAFC, that patentees tended to file cases in more favorable circuits prior to the CAFC, and that both phenomena were mitigated by the CAFC. Henry (forthcoming)7

Appendix E in Turner (2002) includes some of these data, as well as 1982-94 data on validity from Dunner, Jakes and Karceski (1995). Awkwardly, the Federico (1956) and Koenig (1980) data for 1953-54 do not agree. In carefully studying decisions published in the USPQ for 1953-54, Henry and Turner were unable to perfectly replicate either set of numbers. Additionally, Henry and Turner could find no similar archive of USPQ data for 1979 and beyond. Indeed, there were no papers with any patent litigation decision data for 1979-81. As a result of these shortcomings, Henry and Turner developed detailed new criteria for what should be included and set their sights on securing data comprehensively for 1953-2002. These criteria are described in subsection 3.1.
performs an event study on the effect of litigation decisions on firm value. He finds that “invalid”
decisions yield about 0.85% losses for publicly-traded patent-holding firms, and that firms suffer
significantly bigger losses during the CAFC Era. McGahee (2011) compares litigated and matched
patents across a wide array of characteristics. Following Jaffe, Trajtenberg and Henderson (1993),
McGahee and Turner (2013) estimate rates of localization of knowledge spillovers of litigated
patents. We discuss each of these papers in more detail later in the paper.

3. The Data

The data combine two separately-gathered data sets, one for district court decisions published
in the USPQ and another for appeals court decisions published in the USPQ. In both cases,
decisions span 1929-2006. District court decisions are matched to the first appeals court decision
in the case, using the USPQ and case histories from Westlaw. In cases where the district court
decision is not published but the appellate court decision is published, we observe the district
court decision from the appellate court decision. Hence, our data include many district court
decisions that are unpublished.

The USPQ has two main advantages as a data source. First, each published decision includes
the entire written opinion of the case. Unlike alternative sources of data, such as that from the
Federal Judicial Center, one can determine with high accuracy the court’s ruling on the principal
issues of validity and/or infringement for each patent case, for both the district and appellate
court decisions.

Second, the USPQ has published all decisions deemed, by legal editors, to be potentially
precedential or to include noteworthy fact patterns. Virtually every appellate court decision is
recorded in the USPQ, as well as a large sample of district court decisions. Our search yields
recorded opinions from 4,294 unique district court cases and from 4,293 unique appellate court
cases. Among the district court cases with recorded opinions, at least one patent is appealed
to a decision in 2,378. Because cases are sometimes combined on appeal, this results in 2,321
unique appellate court cases. For the remaining 1,972 unique appellate court cases, we observe

---

8We search volumes 1-231 of Edition 1 of the USPQ and volumes 1-80 of Edition 2.
9The USPQ also records some decisions that it deems as “unpublished” which it deems not to have precedential
value. These form around 2.5% of all recorded decisions, and this group is statistically indistinguishable from the
recorded decisions that are “published.”
10This is based on a conversation with managing editor William McKey.
the recorded appellate court opinion but do not observe a district court opinion.

The observational unit in the data set is a *patent case*—a particular patent in a particular case that reached at least a district court decision. For a case with four patents at issue, there are four patent cases. In total there are 9,359 patent cases. We cannot identify the patent number in 29 of these cases. Among the rest, 8,256 are utility patent cases, 592 are reissue patent cases, 460 are design patent cases, and 22 are plant patent cases. Some patents appear in more than one case. Among the 9,330 patent cases where we identify the patent number, there are 8,555 unique patents.

3.1. Standards For Inclusion

Following Henry and Turner (2006), the data are constructed to study how courts have handled the issues of patent validity and infringement. As such, they include only cases that discuss one or both issues. Thus, “housekeeping” decisions, such as improper venue challenges, are excluded. Furthermore, preliminary injunctions are also excluded, because the plaintiff’s burden of proof is higher.\(^{11}\) In addition, cases where the patentee’s conduct is the primary issue (in which a patent might be found “unenforceable”), and cases decided on the basis of collateral estoppel or *res judicata*,\(^{12}\) are excluded.

When a case involves a dispute over a license, it can typically be sorted into one of two categories: (1) the parties have an agreement but the licensee expects to fare better in court than at the bargaining table; or (2) the court finds the patent is not infringed because of how a license is construed. For the former category, decisions are typically based on the patent’s validity or the licensee’s infringement (e.g., the licensee continues to produce without a license). These cases are included. Decisions in the latter class are not based on validity or infringement, so these are excluded.

Many patents are in more than one patent case. When multiple cases are decided in the same circuit, the later case(s) are discarded when the cases are substantially similar. For example, when a court finds a patent “valid and infringed” and, subsequently, the infringing party attempts

\(^{11}\)The statute 35 U.S.C.A. 283 gives discretion to the court to grant preliminary injunctions when it deems it reasonable. The CAFC advocates a four-factor test to determine when granting an injunction is reasonable: 1) a reasonable likelihood of success on the merits; 2) irreparable harm if the injunction were not granted; 3) the balance of the hardships; and 4) the impact of the injunction on the public interest. See, e.g., *Genentech, Inc. v. Novo Nordisk A/S* (108 F.3d 1361, 42 U.S.P.Q 1001 [Fed. Cir. 1997]).

\(^{12}\)The plaintiff can not win a suit if the matter was previously adjudicated and all parties enjoy mutuality.
to invent around the patent, prompting the patentee to sue for infringement again, we exclude the later case. We do include patents adjudicated in different circuits either simultaneously or sequentially, however, because differences in courts’ tendencies across circuits are of significant interest.

3.2. Variables

A sample patent case is shown in Table 1. Full description of variable names and classifications can be found in the codebooks accompanying these data. Patent-specific variables are captured from patent documents archived by the USPTO. Many of our variables are found in typical patent data sets, such as the NBER data, but our set differs in a couple of significant ways. Note that we capture both an Application Date and an Original Application Date. The former is the application date for the final application that was ultimately issued as the patent. The Original Application Date traces the application back, through any divisions or continuations, to the very first application date listed. This distinction more accurately identifies the age of the patented technology and, with data on continuations and divisions, permits careful inquiry into the nature of administrative lags and how they relate to litigation.

Note also that we capture, for patents issued after January 1, 1947, the median-age cited patent and the oldest cited patent. These variables shed light on the age and composition of prior technology. In capturing the median, if there are an even number of backward references, we record the more recent of the two middle references. For the patent in Table 1, for example, the median reference is the third-oldest of six. Intuitively, if the subject patent is included with the references, there are seven patents in the group and this method identifies as the median the patent in the middle of this bigger group.

Litigation variables are captured from the opinions published in the USPQ. Most of the variables are straightforward to classify, but decisions themselves are complicated because courts occasionally judge some of a patent’s claims one way and others another way. Following Federico (1956) and Henry and Turner (2006), we classify such patents as follows. Any patent with at least one “valid and infringed” claim is recorded as “valid and infringed.” Intuitively, the patentee is winning something, as infringement of one claim is sufficient for an award of damages. If there

---

13 The variable names we use in the data files are shortened versions of the more descriptive terms in Table 1.
14 Backward references are not included in patents issued prior to 1947.
are no “valid and infringed” claims, then any patent with at least one “invalid” claim is recorded as “invalid.” Finally, “not infringed” claims trump “not adjudicated” claims (i.e., undecided, remanded, etc.).

Our data also include a matched set of random patents, which may or may not be litigated. To execute this match, we capture all patents sharing at least one primary product code with the subject patent that are issued within one month of the subject patent. From this group, which includes the subject patent, we choose one patent at random to be the control. The rate of self-matching is about 2.1%.

Ideally, we would like to follow Lanjouw and Schankerman (2001) and capture matched patents using just the main product code and using the application date, as this would generate a less noisy control group. There are two problems, however. When just the main product code is used, the rate of self matches is unacceptably high. One would need to use a much wider date-match window to get to 2% self matches. Also, our data include a high number of pre-1976 patents. Because the USPTO archives the issue date for such patents in HTML format (allowing patents to be searched by issue date), but does not archive the application date this way, it would be prohibitively costly to match by application date.

For the patent in Table 1, “Haskett Chlorinating System,” there is only one primary product code, 137, “Fluid Handling.” Issued November 30, 1965, this patent is matched to Patent No. 3,225,974, “Hose Nozzle Supporting Device,” whose product codes are 222, 137, and 242, and whose issue date is December 28, 1965. Our matched-patent data include variables for all patent-specific categories.

4. Basic Descriptive Statistics

The sampling frequency of published district court patent cases changes across time. Table 2 lists the average yearly number of published district court patent cases, the average yearly number of appellate court patent cases, and the ratio of the two. The early years include more observations per year and there is an overall decline in the average. The district/appellate ratio is steady through 1980, with a slight decline through 2000 then a big drop after 2000. The average

---

15 Patents found “valid but not infringed” are recorded as “not infringed.”
16 One would need to first capture application dates from .pdf files of all patents, then run the matching algorithm. Because the USPTO’s formatting differs across patent documents, and because original application dates are typically listed as part of the text of the patent, this would effectively require manual examination of every patent.
number of appellate court patent cases rises sharply after 2000. This is largely a result of the influence of the CAFC. District courts have been less prone to tackle thorny issues of patent law and more inclined to have those issues decided by the CAFC.

There are likely to be applications where researchers will wish to use patent cases in unpublished district court decisions. For example, anyone wishing to study the 1990s and 2000s will dramatically increase their sample size by including such patent cases. In total, about one third of patent cases in our data have unpublished district court decisions.

A natural concern arises as to whether the way published district court decisions are selected by the USPQ is correlated with key variables. If this were the case, selection may make it more difficult to use all of the district court decisions. Fortunately, this selection does not affect rates of validity or infringement, provided changes in court precedent are accounted for. Consider Table 3, which shows rates of validity and infringement for published and unpublished district court patent cases for part of our data. Following the results of Henry and Turner (2013), who identify structural breaks in the rates of validity in District Court decisions for the second quarter of 1939 (when the rate decreases) and the first quarter of 1983 (when the rate increases), we consider rates of validity for published and unpublished during three distinct time periods. For infringement, Henry and Turner (2013) identify breaks for the second quarter of 1950 (the rate increases) and the second quarter of 1991 (the rate decreases). The comparison of published versus unpublished is similarly broken down for these three time periods.\footnote{See the right-hand columns of Tables 1 and 2, Henry and Turner (2013). We discuss specifics of how we estimate validity and infringement in section 5.} The reader may wish to look ahead to Figure 1 (p. 15) to see the dynamics in rates of validity and infringement before considering the comparison in Table 3.

Note that if published decisions are compared to unpublished decisions not conditional on structural breaks, it appears that patent cases in published decisions are less likely to be valid and more likely to be infringed. However, this is an artifact of the changing selection of the USPQ across time (recall Table 2). Once structural breaks in validity and infringement are accounted for, rates of validity and infringement in published decisions virtually indistinguishable from rates of validity and infringement in unpublished decisions. No comparison yields a difference bigger than 4.7%, and none of the differences are statistically significant at the 5% level.
Table 4 reports the number of patent cases by geographical circuit of origin, along with average rates of the three principal outcomes in patent cases. Cases are not distributed evenly among the circuits. While cases are concentrated in the more heavily populated circuits—the Second Circuit includes New York, the Seventh Circuit includes Chicago and the Ninth Circuit includes Los Angeles—population and economic activity do not entirely explain this distribution. For example, the Third Circuit has a large number of cases primarily because Delaware is a popular state of incorporation and hosts an unusually large number of cases. Additionally the Eleventh Circuit was created in 1982, so there are fewer cases originating from its districts.

Next, consider variability in outcomes across circuits. The frequency of invalidity is highest for the Third Circuit and lowest for the Eleventh Circuit. The frequency of patents being found not infringed is highest in the DC Circuit and lowest in the Seventh Circuit.

While it may be appropriate to use all decisions to study rates of patent enforcement across time, the publication selection does affect the study of rates of appeal. All unpublished DC decisions in our data are appealed—it is only through the appellate court decision that we observe the district court decision in the first place. Hence, these must be discarded when studying rates of appeal. Considering only published district court decisions, Table 5 shows the likelihood that a patent case is appealed to a decision. Other than the 1929-40 period, this rate is fairly steady across time. While our data can be used to study differences in rates of appeal across time, types of decisions, circuits, etc., we are hesitant to claim that the estimated levels of these rates are statistically unbiased. Because published district court decisions are characterized by noteworthy fact patterns, it could be that such decisions are more likely to be appealed to a decision.

Finally, note that 652 of our patent cases are part of declaratory judgment actions. In a declaratory judgment, a party that can prove to the court that it has a reasonable apprehension of being sued for patent infringement can instead pre-emptively file suit, asking the court to find patents invalid and/or not infringed. In studying determinants of validity in district court decisions using the 1953-2002 part of this data, Atkinson, Marco and Turner (2009) find patents significantly more likely to be found invalid in declaratory judgments. They also find significant interaction effects between declaratory judgments and other variables, as determinants of the likelihood of validity. Indeed, there are so many interaction effects that Atkinson, Marco and Turner (2009)

---

18There are 164 patent cases decided in courts with no geographic appellate court venue. Venues include the International Trade Commission, US Court of Claims and the US Court of Customs and Patent Appeals.
19The percentages in the table do not all add up to 100% because of rounding.
remove declaratory judgments from the data for much of their analysis. As little is known about why declaratory judgments differ so much from infringement suits, careful study of them using our data would be a novel and useful project for future research.

5. Enforcement Rates Across Time

Denote $\phi_{INV}$, $\phi_{NINF}$ and $\phi_{VI}$ as the rates of “invalid,” “not infringed” and “valid and infringed” decisions. Using our data, each of these rates may be estimated at very fine levels of disaggregation. Henry and Turner (2006), for example, study time-series properties of annual measures of these rates in district court decisions, as well as conditional rates of appeal and rates of appellate court affirmation. Most notably, they show that rates of invalidity and rates of appellate court affirmation of invalidity decisions declined sharply at the advent of the CAFC.

Henry and Turner (2013) develop a novel methodology for estimating rates of enforcement and use the extended 1929-2006 data to study enforcement rates more comprehensively. Let the rates of validity and infringement be $\alpha_V$ and $\alpha_I$. Henry and Turner (2013) specify $\alpha_V = 1 - \phi_{INV}$, and $\alpha_I = \frac{\phi_{VI}}{\phi_{VI} + \phi_{NINF}}$. Namely, to estimate rates of infringement, they effectively omit patent cases where the patent is found invalid. This recognizes that if a patent is found invalid, then the question of infringement is moot. Assuming that the characteristics that determine whether a patent is invalid are statistically independent of the characteristics that determine whether a patent is infringed, then $\alpha_I$ is the overall rate of infringement.\footnote{That is, it would also be the rate at which patents found invalid would be found infringed if the court were to decide infringement in those cases.}

Henry and Turner (2013) use all 78 years of the UGA Patent Litigation Datafile. Once declaratory judgments are removed, there are 8,707 patent cases. Among these, the district court decision date is observed for 6,534. For the other 2,173, the district court decision is not published but the appellate court opinion is published.\footnote{Note that the data do include decision dates for some unpublished district court decisions.} Because appellate court opinions do not typically include the date of the original district court decision, Henry and Turner (2013) estimate the date of the district court decision. The data include both a district court decision and an appellate court decision for 3,650 patent cases. For each appellate court decision year, they estimate the average length of time between the district and appellate decisions. Henry and Turner (2013) then use this estimated average lag to (backwardly) estimate the district court decision date for the 2,173
observations with just an appellate court decision date. Obviously, this increases the number of observations, but there is an additional benefit. As the average lag between district and appellate decisions is about a year and a half, this adds district court decisions estimated to have occurred in 1927 and 1928.

To identify structural breaks, Henry and Turner (2013) adopt the approach of Bai and Perron (1998; 2003). The model is specified

\[ y_t = x_t' \beta + z_t' \delta_j + u_t \quad t = T_{j-1} + 1, \ldots, T_j \]

for \( j = 1, \ldots, l + 1 \). In this model, \( y_t \) is the value of the dependent variable at time \( t \) (for example, the rate of validity in quarter \( t \)), \( x_t \) and \( z_t \) are vectors of covariates, \( \beta \) and \( \delta_j \) are vectors of coefficients and \( u_t \) is the time-\( t \) error. The set of break dates \( \{T_1, \ldots, T_l\} \) is treated as unknown. The vector \( \beta \) does not shift,\(^{22}\) while the \( \delta_j \) vectors may vary with \( j \). Intuitively, intertemporal changes in standards for validity and infringement will be reflected by changes in \( \delta \). The goal is to estimate both the break dates and the regression coefficients.

Because the number of sets of possible break dates expands rapidly as \( l \) increases, estimation of multiple-structural break models is more challenging than single-break models. For one thing, it is algorithmically more complicated just to solve the problem of globally minimizing the sum of squared residuals. Fortunately, Bai and Perron (2003) identify a tractable algorithm, based on dynamic programming, that accomplishes this task. Second, there is some subjectivity to multiple-break estimation because the number of breaks that globally minimizes the sum of squared residuals might not be statistically significant versus an alternative hypothesis of, say, one less break. Bai and Perron (2003) recommend two approaches, information criteria such as the Bayes Information Criteria (BIC) and iterative tests for significance of \( l + 1 \) versus \( l \) breaks, for potentially multiple values of \( l \). Finally, for constructing break-date confidence intervals, Henry and Turner (2013) employ the “shrinking shifts” asymptotic approach.

Henry and Turner (2013) consider three different approaches to estimating validity and infringement. They estimate these rates using just district court decisions (the District Only model), using just appellate court decisions (the Appellate Only model), and using both district and appellate decisions (the Overall model). For each series, they run the Bai-Perron estimation procedure, with

\(^{22}\)Henry and Turner (2013) do not include any variables in \( x_t' \), but this could be done.
a maximum number of five breaks. All series have statistically significant structural breaks, and the breaks are similar across models.

Figure 1 shows rates of validity and infringement from the Overall model, along with fitted lines. For this model, there is sufficient data to construct quarterly rates of validity and infringement. In the validity series, the model identifies breaks in 1939:4 and 1983:1. Patent validity is weakest between 1939 and 1983. During the late 1930s, patents came under severe scrutiny as tools of monopoly, and courts took an increasingly dim opinion of them (Schmookler 1966).

---

23 These results are recorded in the right-hand columns of Tables 5 and 6, Henry and Turner (2013).
24 Numerous hearings of the Temporary National Economic Committee (TNEC) focused on how firms used patents as tools to form cartels. Scherer (2009) notes that one such set of hearings prompted antitrust proceedings...
estimated drop in validity from about 58% to about 40% percent appears to confirm that this had a significant effect on validity outcomes. Estimated validity rates hold steady until the establishment of the CAFC, when they rise sharply to about 74%. Hence, consistent with conventional wisdom, enforcement of patent validity is strongest during the CAFC era.

In contrast, however, patent scope, as interpreted through rates of infringement, appears to be broadest between 1951:3 and 1990:4 and is narrowest since 1990:4. The first structural break occurs closely-timed to the influential May 29, 1950 Supreme Court decision in *Graver Tank & Mfg. Co. v. Linde Air Products Co.* [339 U.S. 605], which established a broad scope of the doctrine of equivalents.\(^{25}\) The second of these structural breaks occurs during a period when the CAFC adopted precedents weakening the doctrine of equivalents, although there is no case of similar influence to *Graver Tank* in the late 1980s or early 1990s.

Interestingly, after the CAFC restricts patent scope, the overall patentee win rate (i.e., the likelihood of winning on both validity and infringement) returns to about 28%, roughly where it was during both before 1939 and between 1951-83. Viewed this way, the early years of CAFC Era, when rates of both validity and infringement were high, appear anomalous. Indeed, for 58 of the 78 years during 1929-2006, the overall win rate is between 27-29%.

6. Uniformity and Forum Shopping

Our data also permit a broad and deep inquiry into the impact of policy changes on the behavior of patentees and courts. Consider the CAFC. Supporters argued that it would unify previously varied interpretations of patent law across the geographical Circuit Courts of Appeal, making validity outcomes in district court decisions more uniform across geographical circuits and mitigating the incentives of litigants to “forum shop” for their preferred venue. These data allow careful statistical tests of these propositions, controlling for patent- and litigation-specific characteristics.

Atkinson, Marco and Turner (2009) use binary choice models to study the impact of the CAFC on uniformity and on forum shopping. For uniformity, the key dependent variable is \(\text{valid}_j\), which

---

\(^{25}\)The doctrine of equivalents states that infringement need not be literal. Rather, if a party uses something that is equivalent to the patented technology, then it may nonetheless infringe. In the *Graver Tank* case, the alleged infringer substituted manganese for magnesium in a welding process, and was held to infringe because manganese was equivalent in that process.
is “1” if the patent is found valid by the district court and is “0” otherwise. The Probit model is specified as follows:

\[ \text{valid}^*_j = \delta + \eta X_{j}^{\text{Era}} + \beta X_{j}^{\text{Trial}} + \theta X_{j}^{\text{Era}} X_{j}^{\text{Trial}} + \varphi Z_j + \varepsilon_j, \tag{1} \]

where valid* is a latent variable measuring the degree to which the patent satisfies the legal requirements for patentability.\textsuperscript{26} The subscript \( j \) indexes the patent cases and \( \varepsilon_j \) is independently and identically distributed Normal with mean zero and constant variance. This error term primarily reflects unobservable factors determining patent validity, such as the competence and performance of randomly assigned judges, special masters and jurors.

The vector \( X_{j}^{\text{Era}} \) is a dummy variable for the CAFC era, \( X_{j}^{\text{Trial}} \) is the vector of trial circuit dummies (districts are aggregated by circuit in both eras), and \( Z_j \) is a vector of controls. Observe:

\[ \text{valid}_j = \begin{cases} 1 & \text{if valid}^*_j \geq 0 \\ 0 & \text{if valid}^*_j < 0 \end{cases}. \tag{2} \]

Conditional on this, maximum likelihood estimates of the parameter vectors \( \beta, \theta \) and \( \beta + \theta \) form the basis of three key tests of uniformity.

**Hypothesis 1.** \( \beta = 0 \). Validity rates are uniform across circuits in the pre-CAFC era, ceteris paribus.

**Hypothesis 2.** \( \theta = 0 \). The effect of the CAFC on validity rates is the same across all circuits, ceteris paribus.

**Hypothesis 3.** \( \beta = -\theta \). Validity rates are uniform across circuits in the CAFC era, ceteris paribus.

Using likelihood ratio tests, Atkinson, Marco and Turner (2009) reject all three null hypotheses. Hence, the data indicate significant non-uniformity in validity rates both eras (H1 and H3) and a significant change in non-uniformity across eras (H2). Since 1982, validity rates have been significantly less dispersed across geographical circuits.

Figure 2 highlights the intuition for these results. Validity rates for the 10 geographic circuits are shown for the pre-CAFC (1953-82) and CAFC (1983-2002) eras.\textsuperscript{27} Each point lies above the

\textsuperscript{26}We regret the different use of \( \beta \) and \( \delta \) in this section versus section 5. We wish to use the same notation as appears in the papers themselves.

\textsuperscript{27}The Fifth and Eleventh Circuits are combined after the latter’s creation in 1982.
45° line, indicating a greater likelihood of validity in the CAFC era for the district courts in each geographical circuit. Also, the variation in validity rates is much larger for the pre-CAFC era. Finally, circuits whose districts were “strong” on validity in the pre-CAFC era maintained those characteristics, somewhat, in the CAFC era, as illustrated by the fitted line.

Supporters of the CAFC argued that by bringing uniformity to district court outcomes, it would mitigate the incentives of patentees to “forum shop” for their preferred interpretation of the law. Atkinson, Marco and Turner (2009) develop, then test, a theory of forum shopping on the basis of patent validity. The model assumes that there is a “natural” forum in which litigation is least expensive, while other fora offer the patentee different probabilities of validity. The model predicts that forum shopping is more likely when probabilities of validity are less uniform across fora, and more likely when the “natural” forum is less favorable on validity.

With data on patentee locations, plaintiff identity, and filing dates, one can conduct careful statistical tests of the presence of forum shopping on the basis of validity, while controlling for key variables. To obtain filing dates, Atkinson, Marco and Turner (2009) match the 1953-2002 part of the UGA data to Federal Judicial Center data.
The key dependent variable is \( \text{away}_j \), which is “1” if the case is litigated outside the patentee’s geographical “home” circuit and is “0” otherwise. Effectively, the “home” circuit is regarded as “natural,” and the key question is whether expected validity, or the presence of the CAFC, affect the likelihood the patentee files its case elsewhere. Specifically, the circuit choice model is as follows:

\[
\text{away}_j = \delta + \eta X_j^{\text{Era}} + \psi X_j^{\text{Era}} \times \text{valid}_j + \varphi Z_j + \varepsilon_j,
\]

where \( \text{away}_j \) is a latent variable measuring the degree to which the patentee prefers to litigate outside the home circuit. The error \( \varepsilon_j \) primarily reflects unobservable random factors, such as infringer location, that affect case location. Then

\[
\text{away}_j = \begin{cases} 
1 & \text{if } \text{away}_j \geq 0 \\
0 & \text{if } \text{away}_j < 0
\end{cases}
\]

If the CAFC did mitigate forum shopping, it is an open empirical question when that mitigation began. To maintain flexibility, \( X_j^{\text{Era}} \) is specified as a set of time dummies from 1963-96, \( \{ \text{yr}63-67, \text{yr}68-72, \text{yr}73-77, \text{yr}78-82, \text{yr}83-87, \text{yr}88-92, \text{yr}93-96 \} \), where the subscript represents a window of filing years.

The variable \( \text{valid}_j \) measures the difference between the recent (five-year) validity rate in the “home” circuit and the aggregate rate, capturing the expected validity advantage of litigating in the “home” circuit. The basic test of forum shopping is whether patentees systematically litigate at home when home district courts have recently been favorable on validity and away from home when home district courts have been unfavorable. Atkinson, Marco and Turner (2009) estimate the effect of \( \text{valid}_j \) on \( \text{away}_j \) and test for its significance. To test for temporal changes in forum shopping, this variable is interacted with the time dummies in \( X_j^{\text{Era}} \), to test the following hypothesis.

**Hypothesis 4.** \( \psi = 0 \). Recent home-circuit validity rates are unassociated with the likelihood the home circuit is chosen as the trial circuit, ceteris paribus.

Atkinson, Marco and Turner (2009) reject this null hypothesis for the 1963-67, 1968-72 and 1973-77 time windows, finding that an increase in the home validity advantage results in an increase of between .05 and .09 in the probability that the case is litigated in the home circuit. For the other
time windows, the null is not rejected. Hence, in the pre-CAFC era, patentees are more likely to litigate in their home circuit when that circuit is more favorable on validity, but forum shopping ceases in 1978. This suggests that patentees anticipated the impact of the CAFC several years prior to its establishment.

Figure 3 highlights the intuition for these results. Define net inflow into the $i$th Circuit is the share of total patent cases litigated in the $i$th Circuit less the share of total patent cases “born” in the $i$th Circuit. If systematic forum shopping (for preferred venues on validity) prevails, then circuits that are weak on validity should export cases to other circuits (so net inflow would tend to be negative). Circuits that are strong on validity should import cases (so net inflow would tend to be positive). Figure 3 shows scatterplots of net inflow vs. validity rates for the pre-CAFC and CAFC eras, along with fitted lines from weighted regressions.

Two features of the figure highlight forum shopping and the impact of the CAFC. First, this figure indicates that case migration (from home to trial circuit) is more concentrated in the pre-CAFC era. The spread in net inflow across circuits (plotted vertically) is much greater in the pre-CAFC era. Circuits 2, 4, 5, 6 and 7 each have, for the pre-CAFC era, net inflow that is larger,
in absolute value, than net inflow for every circuit in the CAFC era.

Second, the relationship between net inflow and validity is consistent with the theory in both eras but far stronger in the pre-CAFC era. The estimated slope coefficient is .16 for the pre-CAFC-era fitted line, but only .03 for the CAFC-era fitted line.\footnote{The former estimate is of marginal statistical significance, while the latter is clearly insignificant.}

With data extended back to 1929 and forward to 2006, it is possible to use models such as (1)-(2) and (3)-(4) to carefully analyze the determinants of invalidity and circuit choice for these additional years. Perhaps most importantly, the new 1929-52 data will permit careful study of the tumultuous 1930s and 1940s.

7. Comparison of Litigated versus Matched Patents

Table 6 shows descriptive statistics for 8,555 litigated and control patents. Although we do capture multiple controls when a patent is litigated more than once, these statistics reflect just one control patent per litigated patent.\footnote{The PatentDup variable identifies which controls are used.} For these statistics, some variables may be calculated in different ways. Here, we say a variable is Assigned if an Assignee is listed on the original patent. We say a patent is a Continuation if the patent describes the application as either a continuation or continuation-in-part. We say a patent is Foreign if the first inventor’s home country is not the United States.\footnote{A handful of patents do not reveal the inventor’s country, so the N is lower for this variable.} The Grant Lag is measured from the date of the original application to the patent’s issue. Because pre-1947 patents do not list backward references, there are fewer observations for the three variables at the bottom of the Table. Because of how we select for controls, the number of controls with backward references is slightly higher than litigated patent with such references.

Standard difference-in-means tests suggest that differences between litigated patents and control patents in our data are similar to differences found by other researchers using more recent data. First, similar to Allison et al. (2004) and Graham, Hedge and Mowery (2009), we find litigated patents are more likely to come from a continuation or division, take longer to issue and have more backward references. Second, consistent with Lanjouw and Schankerman (2001), litigated patents in our data have significantly more claims and are less likely to be of foreign origin. Finally, litigated patents are no more likely to be assigned at issue, have the same number...
of inventors, are equally likely to have at least one backward reference, and have an equally-old median-age backward reference.

In contrast to these other studies, our data permit careful inquiry into intertemporal trends in characteristics of litigated patents. McGahee (2011) studies such trends comprehensively. In this section, we highlight some of his findings. Consider the rate of assignation. Here, we follow McGahee (2011) in referring to assigned patents as “corporate.” Figure 4 shows that the rate of corporate ownership (by patent issue year) increases steadily over time. However, the basic finding that litigated patents are no more likely than controls to be corporate holds across time. Both the litigated and control rates range from about 30 percent prior to 1920 to over 80 percent in the 2000’s, but they move together. Growth in corporate ownership slows in the 1940’s, for litigated and control patents, but picks up in the late 1950’s. McGahee (2011) finds qualitatively similar results for the number of inventors. The mean number of inventors grows from about 1.1 to 2.3 inventors per patent for both litigated and controls. This growth is consistent with the increasing number of corporate patents.

However, intertemporal trends are qualitatively distinct for the frequency of continuations. From Table 6, litigated patents are about twice as likely to issue from continuations as their controls. But Figure 5 shows that prior to about 1960, continuations were very rare. For these earlier years, the difference between litigated patents and controls is small and there are several issue years where more control patents issue from continuations. In the later years, by contrast,
the frequency of continuations grows for both the litigated and the control group, but grows much faster for litigated patents. Over half of litigated patents issued during the 1990s were continuations. The surge appears to begin shortly after the Patent Act of 1952 codified the practice of accepting continuation applications into law.

In a multinomial probit analysis of the determinants of decisions, McGahee controls for a variety of factors and finds no statistically significant difference in the likelihood of a “valid and infringed” outcome (or other outcomes) for litigated patents issued from continuations versus non-continuation patents prior to the Patent Act of 1952. However, the probability of a “valid and infringed” outcome for patents issued from continuations rises by nineteen percentage points more than the concurrent rise for non-continuation patents following its enactment. He also finds that corporate patents were about seven percentage points more likely to be found valid and infringed than non-corporate patents from the very beginning of our data.

8. Matching to Other Data

The variables in our data permit a wide range of possible matches to other data, expanding the scope of its application. Using firm names, it is possible to match the data to company-specific variables in data sets such as COMPUSTAT and CRSP. Using patent numbers, it is straightforward to match these data to the NBER Patent Datafile. This enables linking litigation-
specific variables with the rich forward citation data in the NBER file. Using years, it is possible to match these data to a host of policy events. There are likely to be other useful matches as well. We briefly discuss work underway that takes advantage of other data sets.

Henry (forthcoming) uses a subset of this data from 1963-2002 to study stock market returns of companies for the days around patent litigation decisions. He finds that “invalid” decisions reduce firm value by about 0.8% across the entire time span, but the negative impact is 0.7% larger after the establishment of the CAFC. He argues that since the market expects validity with greater likelihood under the CAFC, an “invalid” decision is a bigger (negative) surprise. This adds additional evidence that the CAFC has been generally viewed as a “pro-patent” court, and helps establish the event-study approach as a way to evaluate the effects of precedent and policy on patent value.

The remainder of the data, representing about half of the years and more than half of the observations, have not been used in event studies as of yet. As a result, we know little about how investors viewed patents between 1929-62. By matching litigation events to stock returns, macroeconomic variables (which primarily reflect determinants of patent value under certain enforcement) and changes to policy and precedent (which primarily affect the level of enforcement), one could use the variability in both macroeconomic conditions and patent enforcement over 1929-2006 to carefully identify patent value.

Given that so many patents are invalidated, it is natural to wonder how many non-litigated patents would be invalidated. This is potentially of great interest for public policy, because the presence of such patents may hinder value-creating activity by increasing the effective cost of using certain technology. Matching the UGA data with company data, Henkel, Turner and Zischka (2013) develop a model geared towards predicting the share of all patents that, if they underwent invalidation proceedings, would be ruled invalid. Early evidence from in-depth interviews with lawyers and judges show that, among several other determinants, firm characteristics might significantly influence invalidation decisions.

Additionally, forward citations of litigated patents may reveal additional information about technology dissemination. McGahee and Turner (2013) obtain case filing dates by matching to the Federal Judicial Center data. They then match patent numbers for patents in our data, issued during 1975-86, to cited patent numbers in the NBER data. Using the approach of Jaffe,
Trajtenberg and Henderson (1993) to identify “controls” for each citing patent, they estimate rates of localization of knowledge spillovers. Inventor-state localization of citations is generally higher for litigated patents than for those studied by Jaffe, Trajtenberg and Henderson (1993). Moreover, inventor-state localization does not fade with age for litigated patents, in contrast to Jaffe, Trajtenberg and Henderson (1993). Perhaps more interestingly, citations display significant trial-state localization—that is, citations to a litigated patent are significantly more likely to be local to the state in which the district court trial occurs than are control patents. Trial-state localization already is present prior to the filing of litigation – and increases with both the onset and conclusion of litigation.

Finally, invalidated patents are technically worth nothing, yet they do receive citations. No systematic research has studied how citations respond to litigation decisions. Our data are ideal for such an inquiry.

9. Conclusion

We hope these data prove useful to other researchers in their work. While we cannot guarantee the data are perfectly recorded, we have (in drafting our papers) caught and cleaned up numerous errors in the data since completing the major parts of the data collection, and we are confident that the remaining noise should not be severe. We encourage researchers to please communicate with us about ways to improve or expand the data. In addition to the data archived at UGA, we have collected incomplete information on attorneys, judges and some other variables. We hope to add these variables to our archive some time in the future.
REFERENCES


Small Firms Handicapped?,” *Journal of Law and Economics* 47: 45-74.


Table 1: A Sample Patent Case
<table>
<thead>
<tr>
<th>Years</th>
<th>Published DC Decisions Per Year</th>
<th>Total AC Decisions Per Year</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1929-1940</td>
<td>151.0</td>
<td>136.3</td>
<td>1.11</td>
</tr>
<tr>
<td>1941-1950</td>
<td>102.8</td>
<td>81.0</td>
<td>1.27</td>
</tr>
<tr>
<td>1951-1960</td>
<td>82.5</td>
<td>65.0</td>
<td>1.27</td>
</tr>
<tr>
<td>1961-1970</td>
<td>85.6</td>
<td>70.9</td>
<td>1.21</td>
</tr>
<tr>
<td>1971-1980</td>
<td>67.0</td>
<td>53.4</td>
<td>1.25</td>
</tr>
<tr>
<td>1981-1990</td>
<td>63.0</td>
<td>64.4</td>
<td>0.98</td>
</tr>
<tr>
<td>1991-2000</td>
<td>45.6</td>
<td>65.7</td>
<td>0.69</td>
</tr>
<tr>
<td>2001-2006</td>
<td>14.2</td>
<td>109.2</td>
<td>0.13</td>
</tr>
<tr>
<td>All</td>
<td>81.6</td>
<td>80.7</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Table 2: *Average Numbers of Decisions, 1929-2006*

*Note:* These statistics reflect all published district court decisions and all appellate court decisions in our data.
<table>
<thead>
<tr>
<th></th>
<th>Published DC Decisions</th>
<th>Rate (%)</th>
<th>Unpublished DC Decisions</th>
<th>Rate (%)</th>
<th>Difference In Means (Z-Stat)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Validity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-1939:3</td>
<td>1,520</td>
<td>63.5%</td>
<td>942</td>
<td>63.6%</td>
<td>0.59</td>
</tr>
<tr>
<td>1939:3-1983:1</td>
<td>3,442</td>
<td>45.1%</td>
<td>822</td>
<td>48.9%</td>
<td>1.94</td>
</tr>
<tr>
<td>Post-1983:1</td>
<td>921</td>
<td>73.4%</td>
<td>1,060</td>
<td>72.6%</td>
<td>-0.38</td>
</tr>
<tr>
<td>Total</td>
<td>5,883</td>
<td>54.3%</td>
<td>2,824</td>
<td>63.1%</td>
<td>7.84</td>
</tr>
<tr>
<td><strong>Infringement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-1950:3</td>
<td>1,493</td>
<td>61.2%</td>
<td>729</td>
<td>61.2%</td>
<td>0.02</td>
</tr>
<tr>
<td>1950:3-1991:2</td>
<td>1,355</td>
<td>70.5%</td>
<td>410</td>
<td>69.0%</td>
<td>0.56</td>
</tr>
<tr>
<td>Post-1991:2</td>
<td>347</td>
<td>38.3%</td>
<td>642</td>
<td>43.0%</td>
<td>-1.43</td>
</tr>
<tr>
<td>Total</td>
<td>3,195</td>
<td>62.7%</td>
<td>1,781</td>
<td>56.4%</td>
<td>4.29</td>
</tr>
</tbody>
</table>

Table 3: *Comparing Published District Court Decisions to Unpublished District Court Decisions*

**Note:** These statistics reflect all district court decision in infringement suits (declaratory judgments are removed). The choice for time periods reflects structural break analysis of quarterly time series of rates of validity and infringement, using the same data (Henry and Turner 2013, Tables 1 and 2, far right columns).
<table>
<thead>
<tr>
<th>Circuit</th>
<th>N</th>
<th>Invalid</th>
<th>Not Infringed</th>
<th>Valid and Infringed</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>418</td>
<td>49.0%</td>
<td>24.2%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Second</td>
<td>1,675</td>
<td>42.4%</td>
<td>23.2%</td>
<td>34.3%</td>
</tr>
<tr>
<td>Third</td>
<td>1,122</td>
<td>51.2%</td>
<td>21.4%</td>
<td>27.4%</td>
</tr>
<tr>
<td>Fourth</td>
<td>565</td>
<td>39.8%</td>
<td>20.4%</td>
<td>39.8%</td>
</tr>
<tr>
<td>Fifth</td>
<td>606</td>
<td>34.0%</td>
<td>22.8%</td>
<td>43.2%</td>
</tr>
<tr>
<td>Sixth</td>
<td>1,183</td>
<td>46.6%</td>
<td>22.0%</td>
<td>31.4%</td>
</tr>
<tr>
<td>Seventh</td>
<td>1,448</td>
<td>47.6%</td>
<td>16.6%</td>
<td>35.8%</td>
</tr>
<tr>
<td>Eighth</td>
<td>543</td>
<td>41.8%</td>
<td>25.6%</td>
<td>32.6%</td>
</tr>
<tr>
<td>Ninth</td>
<td>1,226</td>
<td>40.9%</td>
<td>27.2%</td>
<td>31.8%</td>
</tr>
<tr>
<td>Tenth</td>
<td>269</td>
<td>36.1%</td>
<td>21.6%</td>
<td>42.4%</td>
</tr>
<tr>
<td>Eleventh</td>
<td>108</td>
<td>27.8%</td>
<td>27.8%</td>
<td>44.4%</td>
</tr>
<tr>
<td>DC</td>
<td>32</td>
<td>43.8%</td>
<td>31.3%</td>
<td>25.0%</td>
</tr>
<tr>
<td>Other</td>
<td>164</td>
<td>43.9%</td>
<td>26.8%</td>
<td>29.3%</td>
</tr>
<tr>
<td>Total</td>
<td>9,359</td>
<td>43.9%</td>
<td>22.4%</td>
<td>33.7%</td>
</tr>
</tbody>
</table>

Table 4: Probabilities of District Court Decisions, By Circuit, 1929-2006

Note: These statistics reflect all patent cases in the data.
<table>
<thead>
<tr>
<th>Years</th>
<th>Published DC Decisions</th>
<th>Decisions Appealed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1929-1940</td>
<td>1,812</td>
<td>43.0%</td>
</tr>
<tr>
<td>1941-1950</td>
<td>1,028</td>
<td>52.9%</td>
</tr>
<tr>
<td>1951-1960</td>
<td>825</td>
<td>55.8%</td>
</tr>
<tr>
<td>1961-1970</td>
<td>856</td>
<td>57.1%</td>
</tr>
<tr>
<td>1971-1980</td>
<td>670</td>
<td>53.1%</td>
</tr>
<tr>
<td>1981-1990</td>
<td>630</td>
<td>59.2%</td>
</tr>
<tr>
<td>1991-2000</td>
<td>456</td>
<td>59.2%</td>
</tr>
<tr>
<td>2001-2006</td>
<td>85</td>
<td>56.5%</td>
</tr>
<tr>
<td>All</td>
<td>6,362</td>
<td>52.2%</td>
</tr>
</tbody>
</table>

Table 5: Rates of Appeal, 1929-2006

Note: These statistics reflect all patent cases in published district court decisions in the data. For these patent cases, if there is a USPQ citation to an appellate court decision in the patent case (even if the case is dismissed), then we count the case as appealed.
<table>
<thead>
<tr>
<th></th>
<th>Litigated</th>
<th></th>
<th>Control</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Assigned</td>
<td>0.63</td>
<td>0.48</td>
<td>8,555</td>
<td>0.62</td>
<td>0.48</td>
<td>8,555</td>
</tr>
<tr>
<td>Claims</td>
<td>12.60</td>
<td>16.74</td>
<td>8,555</td>
<td>8.08</td>
<td>8.23</td>
<td>8,555</td>
</tr>
<tr>
<td>Continuation</td>
<td>0.14</td>
<td>0.35</td>
<td>8,555</td>
<td>0.07</td>
<td>0.25</td>
<td>8,555</td>
</tr>
<tr>
<td>Division</td>
<td>0.07</td>
<td>0.25</td>
<td>8,555</td>
<td>0.04</td>
<td>0.19</td>
<td>8,555</td>
</tr>
<tr>
<td>Foreign</td>
<td>0.08</td>
<td>0.27</td>
<td>8,552</td>
<td>0.19</td>
<td>0.39</td>
<td>8,549</td>
</tr>
<tr>
<td>Grant Lag</td>
<td>3.47</td>
<td>2.80</td>
<td>8,555</td>
<td>2.71</td>
<td>1.81</td>
<td>8,555</td>
</tr>
<tr>
<td>Number of Inventors</td>
<td>1.37</td>
<td>0.81</td>
<td>8,555</td>
<td>1.37</td>
<td>0.85</td>
<td>8,555</td>
</tr>
<tr>
<td>Backward References</td>
<td>10.30</td>
<td>17.47</td>
<td>4,607</td>
<td>7.22</td>
<td>7.96</td>
<td>4,612</td>
</tr>
<tr>
<td>Median Reference Age</td>
<td>12.97</td>
<td>10.71</td>
<td>4,486</td>
<td>13.16</td>
<td>11.90</td>
<td>4,487</td>
</tr>
<tr>
<td>Zero References</td>
<td>0.03</td>
<td>0.16</td>
<td>4,607</td>
<td>0.03</td>
<td>0.16</td>
<td>4,612</td>
</tr>
</tbody>
</table>

Table 6: Litigated Versus Random Patents

**Note:** These statistics reflect 8,555 patents involved in validity or infringement decisions published in the *United States Patents Quarterly* between 1929 and 2006 and 8,555 control patents matched to those litigated patents. Controls are matched within one month of issue date and with at least one common USPC classification. Patents issued prior to 1947 do not include a list of references. Some patents do not include information about the inventor’s country of origin. This Table is very similar to Table 1 from McGahee (2011). Differences reflect slight changes to the data and different techniques for classifying variables (especially Foreign).