

OPENING PANDORA'S BOX: ANALYZING THE COMPLEXITY OF U.S. PATENT LITIGATION

Jonathan H. Ashtor*

18 YALE J.L. & TECH. 217 (2016)

ABSTRACT

Patent litigation is widely regarded as one of the most complex types of civil litigation, with costs often totaling millions of dollars and typical cases lasting years. Also, the burdens of patent case complexity land on both sides of the technological divide, as large producers face skyrocketing defense budgets and inventors and startups risk being “priced out” from enforcing their rights. Yet, the complexity of patent cases is poorly understood as an empirical matter. Instead, patent litigation is generally accepted to be a “Pandora’s Box” of incalculable complexity, which, once opened, is only arduously and unpredictably concluded.

This study undertakes a comprehensive exploration of patent litigation complexity, first defining robust metrics of complexity and continuing with rigorous analysis of the determinants thereof. We focus our study on the eight years of U.S. District Court litigation leading up to passage of the America Invents Act, and we mine extensive detail of more than 1000 cases during this timeframe. Using these data we ask targeted questions about patent case complexity, including what types of cases are most complex, how defense costs compare to enforcement costs, what factors are associated with particularly high complexity, and how complexity has changed over time.

* Associate, Skadden, Arps, Slate, Meagher & Flom LLP; Thomas Edison Innovation Fellow, Center for the Protection of Intellectual Property (CPIP) at George Mason University School of Law. The research and writing of this work was supported by a Leonardo Da Vinci Research Grant granted by CPIP. Any views expressed herein are solely those of the author and do not reflect the views of others, including PricewaterhouseCoopers LLP, Skadden, Arps, Slate, Meagher & Flom LLP (or its attorneys or clients), CPIP, or any of their respective affiliates.

The author is very grateful for the insightful suggestions, comments and feedback of Jay Kesan through multiple rounds of workshops as part of the Thomas Edison Innovation Fellowship, as well as the instructive input of other senior commentators and fellows, including (in alphabetical order) John Duffy, Stu Graham and Zorina Kahn (senior commentators), and Kirti Gupta, Deepak Hegde, Chris Holman, Ryan Holte, Camilla Hrdy, Kristen Osenga, Yi Qian, Ted Sichelman and Saurabh Vishnubhakat (co-fellows). The participants of WIPIP 2016 are also thanked. Finally, utmost gratitude to Adam Mossoff and Mark Schultz, co-founders of CPIP, whose support, encouragement, and intellectual guidance made this work possible.

Finally, we conduct a large-scale event study to identify the causal impact of key policy changes on case complexity, specifically the landmark shifts in remedies law over a series of recent Federal Circuit decisions.

The analysis herein is of crucial importance to patent policy. As juridical property, patent rights are ultimately enforced and defended against in legal proceedings, and thus the complexity of such proceedings directly impacts the rights afforded by patents and recourse thereunder. Understanding case complexity is therefore a necessary contribution to patent policy discourse. Moreover, the framework developed herein sets the stage for future analysis of the complexity impact of new policy measures.

TABLE OF CONTENTS

I.	INTRODUCTION	218
II.	THEORETICAL BACKGROUND	220
	A. Theoretical Framework.....	220
	B. Empirical Scholarship—Litigation Process and Outcomes	222
	C. Empirical Scholarship—Likelihood of Litigation	224
III.	DATASET AND METHODOLOGY	224
	A. Construction of Complexity Metrics	225
	B. Distributions of Case Complexity	227
	C. Analysis of Complexity Metrics.....	230
IV.	EMPIRICAL ANALYSIS	235
	A. Complexity by Phase of Litigation.....	235
	B. Regression Analysis	238
	C. Time Trend Analysis and Event Study	244
V.	CONCLUSION	253
VI.	APPENDICES	256

I. INTRODUCTION

Patent litigation complexity is fundamental to patent rights, implicating both the economic value and burdens thereof. Economic theory explains that efficient outcomes are likely in the absence of significant transaction costs—that is, patents will be licensed at appropriate prices, bad patents will be summarily invalidated, and injunctions of appropriate scope

will issue, in situations where these outcomes are merited. High litigation complexity impedes this process by skewing incentives on both sides. Patent holders are motivated to seek broader remedies for more speculative claims, knowing that accused infringers would rather settle than pay higher costs in litigation. In turn, accused infringers have incentives to “hold up” patent holders who bring legitimate claims, knowing that they can drive down the price of settlement or simply refuse to take a license when the patent holder cannot afford to litigate a case to judgment.

Litigation complexity also has fundamental normative implications for patent rights. Patents are juridical property—as such, the scope and strength of the rights they confer are determined by, and dependent on, the legal environment in which they are enforced. Furthermore, the inventions claimed by patents are not otherwise “excludable,” such that the only recourse to prevent infringement is via judicial order. Accordingly, it holds true for patents that justice delayed is often justice denied, and therefore the causes and characteristics of litigation complexity merit attention.

Nonetheless, with notable exceptions, patent litigation complexity largely remains a black box in legal scholarship. Data on litigation costs and attorneys’ fees are not generally available, and study by way of proxies therefore remains limited. Accordingly, although “everyone knows” that patent cases are highly complex and costly, nobody really know *how much* so, or *why*.

In this study, we analyze over one thousand fully-litigated patent cases in U.S. District Courts that were concluded prior to passage of the America Invents Act (AIA). For each case, we gather detailed information on the initial claims, final dispositions, and the full litigation process in between, examining the recorded dockets and constructing metrics to measure the complexity of the proceedings. We construct a dataset of over 150 unique variables for each case, describing the asserted patents, characteristics of the litigants, and procedural posture and outcome of the disputes. Using a range of statistical techniques, we investigate case complexity and the factors associated therewith, including specific analyses of how complexity varies by type of disposition, phase of litigation, party size and industry characteristics, and key patent attributes.

Finally, we analyze the changes in patent litigation complexity over time, finding a significant increase in the years leading up to the AIA. However, this change is not uniformly distributed across all patent cases, or all phases of litigation, and we trace the principal sources of this increase to the discovery and claim construction phase in cases where the accused infringer prevails. Conversely, we also observe a sharp

increase in the complexity of trials in cases where the patent holder wins. To investigate the underlying causes, we conduct a statistical event study to determine whether the increased complexity of patent trials results from recent changes in remedies case law, particularly the standards for assessing damages. We find strong evidence supporting this hypothesis, and we discuss implications below.

The framework developed herein further sets the stage for analyzing the impact that other patent policy measures have had or are likely to have on litigation complexity. Most notably, the AIA instituted several significant changes to the patent litigation process, including post-grant patent review and amendments to joinder and venue rules. The results of the present analysis offer a baseline against which to evaluate the impact of the AIA, and the empirical metrics and methodology constructed herein are useful to study post-AIA patent cases, once sufficient years of data become available.

This study is organized as follows. In Part II, we outline relevant theoretical background and prior scholarship concerning litigation dynamics generally and patent litigation specifically. Next, Part III describes our dataset and empirical methodology. Part IV provides the various analyses and presents quantitative results. Interpretations and conclusions follow in Part V.

II. THEORETICAL BACKGROUND

Several different avenues of scholarship inform the present study. From a theoretical perspective we draw on the work of Priest and Klein, Eisenberg, and others in understanding the dynamics between and incentives driving parties in litigation. Empirically, Kesan and Ball's work analyzing patent litigation pendency and disposition provides a foundational framework for our methodology. Also, Lanjouw and Schankerman's analysis of the likelihood of patent cases' being litigated, as well as further work in this vein by Chien and Kesan, Schwartz and Sichelman, respectively, explain the selection effects that determine the composition of our dataset. Finally, we also cite to survey studies by the AIPLA and other work to inform our questions about enforcement costs and our expected results.

A. Theoretical Framework

In *The Selection of Disputes for Litigation*, Priest and Klein examine the relationship between fully litigated cases

and disputes that are settled before final adjudication.¹ Based on first principles that settlement is less expensive than litigating and rational parties would seek to minimize their risk and optimize their outcomes, they conclude that disputes should only proceed to trial at the margins—where each party estimates their chances of success to be approximately 50 percent (*i.e.*, no discernible advantage between the parties). Accordingly, they expect that the outcomes of these disputes should be randomly distributed around a mean of 50 percent success for each side, meaning plaintiffs and defendants should each win approximately half the time when they litigate to adjudication. They further test this hypothesis empirical with litigation data from Cook County, Illinois, finding a 50 percent success rate in most types of cases, and they explain the few categories that depart from this expectation.² Their work demonstrates that litigated cases are not a representative sample of all disputes, and therefore conclusions drawn from litigation may be inapplicable to the population as a whole (more on this below).

Importantly, they also note one exception that is expected to produce deviations from the 50 percent norm. Where stakes are particularly high on one side (such as where a dispute implicates other assets or businesses of a party), they predict that one or both parties will be unwilling to settle and even those with clear odds of success *ex ante* may proceed to trial.³ Notably, patent litigation is one of the principal areas which does not conform to the Priest-Klein 50% hypothesis, and many studies have reported patent holder success rates of approximately 30 percent in fully litigated cases.⁴

Notably, Eisenberg tested the 50 percent hypothesis and found that it is not necessarily a general rule for all civil litigation.⁵ Best suited to tort litigation, the 50 percent rule is actually a unique result derived from the selection of tort cases for trial.⁶ However, in other types of cases, working from the same principles of what incentives govern settlement behavior, the sample of disputes expected to be litigated should yield a different success rate. For example, Eisenberg observes a 38 percent success rate in medical malpractice cases in federal

¹ George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 5 (1984).

² *Id.* at 31.

³ *Id.*

⁴ See Rantanen, Jason, *Why Priest-Klein Cannot Apply to Individual Issues in Patent Cases* (U Iowa Legal Studies Research Paper No. 12-15, 2013), <https://perma.cc/4V66-KT2T> (discussing proponents and critics of the Priest-Klein hypothesis, particularly as applied to patent cases).

⁵ Theodore Eisenberg, *Testing the Selection Effect: A New Theoretical Framework with Empirical Tests*, 19 J. LEGAL STUD. 337, 339 (1990).

⁶ *Id.* at 357.

court, which is consistent with his refined theorization of the selection effects in this particular type of litigation.⁷

It is crucial to note that Priest-Klein and subsequent work were aimed at refuting literature that sought to rely on litigated cases as a representative sample of all disputes. By demonstrating the specific selection of which disputes are likely to be litigated and which are likely to settle, they demonstrated that litigated cases are not necessarily representative of the population of disputes but are more likely, in fact, to be outliers. This has central importance for the present study, because we are focusing on fully litigated cases and seeking to understand their behavior and draw policy conclusions therefrom. Moreover, the vast majority of patent cases settle before adjudication—at least on the order of 80 percent⁸—and therefore our dataset represents the “tip of the iceberg” of all patent disputes.

Although we cannot infer that the complexity of cases that are fully litigated has any bearing on the complexity of disputes that settle prior to adjudication (or assertion), we contend that the “tip of the iceberg” is likely to influence party behavior across a range of patent disputes and transactions. That is, the high costs and complexity of patent trials is a strong factor in parties’ decisions to avoid litigation and negotiate settlement. Additionally, where the parties cannot or refuse to reach agreement, adjudication is the only means available to enforce a patent or for the accused infringer to achieve freedom to operate. Thus, litigated cases are precisely those cases that matter most to giving effect to the rights secured by patents—the right of the patent holder to exclusive practice without infringement and the freedom of everyone else to practice around the limits of patent claims without restriction.

B. Empirical Scholarship—Litigation Process and Outcomes

In *How Are Patent Cases Resolved? An Empirical Examination of the Adjudication and Settlement of Patent Disputes*,⁹ Kesan and Ball undertake an unprecedented large-scale empirical study of patent lawsuits to examine rates of settlement, pendency until adjudication or settlement, and case outcomes and specific holdings. By examining docket records

⁷ *Id.*

⁸ See Jay P. Kesan & Gwendolyn G. Ball, *How Are Patent Cases Resolved? An Empirical Examination of the Adjudication and Settlement of Patent Disputes*, 84 WASH. U. L. REV. 237, 259 (2006).

⁹ *Id.*

of each case filed across three annual cohorts, they identify and precisely catalogue the disposition of each case and provide a detailed view of patent adjudication in the District Court system. In particular, they determine the actual rates of settlement of patent cases to be approximately 80 percent, rather than the commonly assumed rate of 95 percent.¹⁰ Also, they identify summary judgment as a principal mechanism for adjudicating patent cases on their merits, and they further find that although summary judgments occur earlier in proceedings than trial, they are not likely to be substantially less costly, based on their measures of litigation expenditure.¹¹

Importantly, Kesan and Ball pioneer a new metric for studying litigation expenditures, focusing on the number of documents filed by all the parties in each case, rather than the time elapsed from complaint to disposition. They explain that case duration is a poor proxy of litigation costs as it “is notoriously inaccurate due to the idiosyncrasies of court schedules and the like,”¹² whereas docket entries are likely to be “more closely correlated with actual litigation costs.”¹³ We follow their approach, and we further build upon it to analyze relative party expenditures in a given litigation.

Another notable study has employed surveys to estimate average litigation costs in dollar values. The most widely cited figures derive from survey studies by the American Intellectual Property Law Association (AIPLA), which estimate that legal costs can range from \$500,000 to \$3 million per suit, or \$500,000 per claim at issue, for each party involved in the litigation.¹⁴ However, specific case-by-case data on fees and litigation costs is not generally available, and average values do not provide meaningful guidance because actual costs are highly skewed and vary widely between cases.¹⁵

Moreover, even if actual fee data were available in every case, it would be unlikely to accurately represent the full costs of the litigation. Attorneys' fees do not account for the internal

¹⁰ *See id.* at 264.

¹¹ *See id.* at 246.

¹² *Id.* at 258.

¹³ *Id.* at 311.

¹⁴ *See* AM. INTELL. PROP. LAW ASS'N, REPORT OF THE ECONOMIC SURVEY (2015).

¹⁵ Work concerning the recent fee-shifting debates has used court awards of the losing party's fees in those cases where fee-shifting applied. *See, e.g.*, James Bessen & Michael J. Meurer, *The Private Costs of Patent Litigation*, 9 J.L. ECON. & POL'Y 59, 80–82 (2012); *see also* Saurabh Vishnubhakat, *What Patent Attorney Fee Awards Really Look Like*, 63 DUKE L.J. ONLINE 15 (2014) (analyzing fee shifting decisions in view of debates over reforms to the standards for fee-shifting in patent cases). However, these datasets do not include a sufficient sample of cases to allow systematic study of litigation costs.

resources, time, and effort devoted by each party to the case, particularly in handling discovery, analyzing technical issues, strategizing, and, in some instances, designing around infringement. Accordingly, although proxies for litigation expenditures do not provide precise dollar values of actual costs, they are the best measure available. We proceed by using various proxies for overall litigation complexity, in the attempt to provide a representative and unbiased assessment of the total expenditure that patent cases entail.

C. Empirical Scholarship—Likelihood of Litigation

Also relevant to our present focus is research on the factors that give rise to patent case filings in the first instance. The work of Lanjouw and Schankerman is formative in this area, particularly their studies that examine the predictors of patent infringement suits in a broader economic context.¹⁶ For example, they examine market and industry factors, litigant characteristics, patent densities and technology fields, and they investigate correlations of these factors with case filings. These studies identify certain characteristics of parties and patents that increase the likelihood of a suit being filed in a particular market/industry and competitive dynamic. For example, they find that the probability of patent litigation increases with respect to patents that are central to follow-on innovations of a company, particularly between companies that are close rivals or where the patent holder needs to maintain a reputation for aggressive enforcement.¹⁷ By contrast, companies in concentrated industries or with particularly large patent portfolios relative to others are less likely to engage in litigation as they often can engage in cross-licensing or have other means of avoiding disputes.¹⁸

III. DATASET AND METHODOLOGY

We begin our analysis of patent litigation complexity by constructing a comprehensive dataset of cases decided in U.S. District Courts during the eight years preceding enactment of

-
- ¹⁶ Jean O. Lanjouw & Mark Schankerman, *Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators*, 114 *ECON. J.* 441 (2004); Jean O. Lanjouw & Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, 47 *J.L. & ECON.* 45 (2004); Jean O. Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 *RAND J. ECON.* 129 (2001).
- ¹⁷ Lanjouw and Schankerman, *Characteristics of Patent Litigation*, *supra* note 15, at 129-30.
- ¹⁸ Lanjouw and Schankerman, *Protecting Intellectual Property Rights*, *supra* note 15, at 48.

the America Invents Act (“AIA”). Our dataset includes over 1000 cases decided between 2004 and 2011, and we undertake extensive individual research of the litigation proceedings.

Specifically, with the assistance of a team of researchers, we parse the full dockets of each case and identify each motion of the patent holder and accused infringer and every substantive court order during the litigation. We further read the complaints and final dispositive orders and catalogue the types of claims asserted in each case, final outcome and specific holdings regarding each of the patents at issue, and we flag the presence of certain key events, such as *Markman* hearings (which construe the claims of the patents at issue), injunctions and appeals.¹⁹

Finally, we incorporate detailed contextual, party and patent information, which facilitates the multivariate statistical analysis below. We conduct background research on the litigants to identify their primary industries and size, whether they are publicly traded or privately held, and whether the patent holder in each case is a practicing entity or non-practicing entity (NPE). Finally, we identify each patent asserted and code an extensive array of attributes using commercial patent databases. Appendix A provides additional description of our dataset composition and Appendix B describes our variables, as resources for other researchers.

A. Construction of Complexity Metrics

We construct four quantitative metrics to evaluate and analyze patent litigation complexity, as follows. Each metric represents increasing complexity, such that a higher metric value represents a more complex case according to that metric.

- **Case Duration:** For each case we analyze the dockets and read the initial complaints and final dispositive orders by the district court to identify the dates when the case was initiated and concluded at the district-court level, respectively.²⁰ We also record the date of the final opinion, order or trial at the district-court level; for example, in cases resolved by summary judgment we code the date of

¹⁹ This work entailed a tremendous research effort by many people, including (in alphabetical order) Daniella Carelli, Courtney Daukas, Josh Glazer, Grace Haidar, Erika Szmanski and Devin Wright, among others. The author is very grateful for their time, effort and perseverance. Special thanks also for the thoughtful insights and contributions of Amber Will.

²⁰ Where there were multiple amended complaints we used the original complaint, even in cases where this was originally filed several years prior to disposition, in order to provide the most comprehensive timeframe for each dispute.

the court's summary judgment ruling, whereas in cases involving trials we record the date when the jury delivered the verdict (if no subsequent court order was issued) or otherwise when the court issued the final order based on the verdict.

- Total Docket Entries and Docket Entries to Disposition: We also code the total number of docket entries involved in the district-court proceedings, with the aim of filtering out noise from various circumstances that can prolong or shorten the duration of individual cases. This approach was pioneered by Kesan and Ball,²¹ and we further extend their methodology by counting both the total number of entries in the district-court docket as well as the entry index number corresponding to the final disposition. This allows us to distinguish and analyze the post-judgment phase of litigation, which may include notices of appeal, motions for fee-shifting, remittitur or vacatur of damages awards and other continuing proceedings.
- Party Motions and Court Orders:²² Finally, to enable analysis of the relative effort and expenditure of patent holders relative to accused infringers, we read each docket and hand-count the number of substantive party motions and court orders between the initial complaint and final disposition.²³ To reduce noise, we count only the initial motion filed by a party, excluding additional related filings—for example, where adjacent entries were labeled as a brief or certificate of service, we counted only the motion itself. We also identify the name of the party filing each motion and code whether it is one of the patent holders or accused infringers in the case. Finally, we identify each instance of the word “order” appearing in the docket and read the title of the resulting entry to determine if it is a substantive court order. We excluded docket entries such as “transcripts” or “minutes” of proceedings, “scheduling orders,” etc. As with party motions, where entries adjacent to a court order were related to it, such as memoranda and opinions regarding a

²¹ Kesan and Ball, *supra* note 7.

²² These metrics were constructed for cases from 2004-2009.

²³ We counted the original complaint and answer as a “motion” and otherwise searched for the word “motion” in the docket file and read the title of the entry to ensure it was appropriately classified as a motion (rather than, for example, a “brief in support of *motion*”).

short-form order, we counted the group only once to avoid redundant entries.

Notably, responsive filings of a party to a given motion are excluded from our counts, as this would eliminate the specificity regarding which party initiated the motion in question. Thus, one would expect the number of court orders to roughly equal the sum of each party's individual motions, on average, given that typically the court will need to issue a ruling on each substantive motion filed by either party. This result is illustrated in the charts presented below.

B. Distributions of Case Complexity

Here we report principal statistics regarding each of our complexity metrics, namely case duration (complaint to disposition by the District Court), docket entries to disposition and substantive docket entries (party motions and court orders). Table 1 below reports the mean, median, and standard deviation of each of our complexity metrics.

Statistic	Mean	Median	Std. Deviation
<i>Case Duration (days):</i>	1003	822	714
<i>Docket Entries to Disposition:</i>	283	205	284
<i>Subst. Docket Entries:</i>	72	53	75

Table 1: Statistics of Overall Complexity Metrics

Based on this data, the average duration of patent cases is over 2.7 years from complaint to disposition via trial or summary judgment (see first row above). Duration also varies substantially across cases, with a standard deviation on the order of 2 years (see above). The litigation process entails considerable activity as recorded in the dockets, with close to 300 entries on average, approximately 70 of which are full substantive motions and court orders (third row in table above). Both of these measures also vary widely, with standard deviations approximately equal to the respective means.

Notably, there is a substantial difference between overall docket entries and substantive docket entries, reflecting the fact that single substantive events in a litigation often give rise to multiple docket entries. For example, a motion for summary judgment could involve a large number of briefs, supporting affidavits and exhibits, supplements and

amendments, etc. Although each of these entries requires some expenditure of substantial cost and effort on the part of the litigants and/or the court, the actual amount thereof is likely to vary considerably. For example, significantly more costs and effort are associated with a motion for summary judgment than a motion *in limine* to exclude certain evidence, although even the latter requires attorney time and incurs related expenses. Filtering for substantive motions and orders thus allows us to focus on relevant litigation events that contribute to overall complexity, excluding non-substantive, interrelated, and duplicative docket entries.

We illustrate the distributions of each metric in Figure 1 (Histogram of Case Duration), Figure 2 (Histogram of Docket Entries to Disposition), and Figure 3 (Histogram of Substantive Docket Entries). These figures illustrate the Poisson characteristics of the distributions, which are expected given the positive whole integer nature of the duration or docket counts in question.²⁴

²⁴ We log-transformed each metric to approximate a normal distribution for linear regression analysis, and we employed statistical tests to ensure that transformed metrics were suitably normal for standard modeling and significance testing. Specifically, we performed Kolmogorov-Smirnov (K-S) tests to compare the log-transformed distribution against a randomly-generated normal distribution having the same mean and standard deviation. We generated ten-fold repeated random normal samples having the same mean and standard deviation as the log-transformed distribution, and we averaged the K-S test results over these 10 iterations. As shown below, the results confirm that the log-transformed metrics are normally distributed (p-values are large, supporting the null hypothesis of identity):

	<u>Ln. Case Duration</u>	<u>Ln. Tot. Entries</u>	<u>Ln. Entries to Disposition</u>	<u>Ln. Subst. Entries</u>
d-Val:	0.0537	0.0557	0.0425	0.0615
p-Val:	0.1447	0.1835	0.3943	0.2584

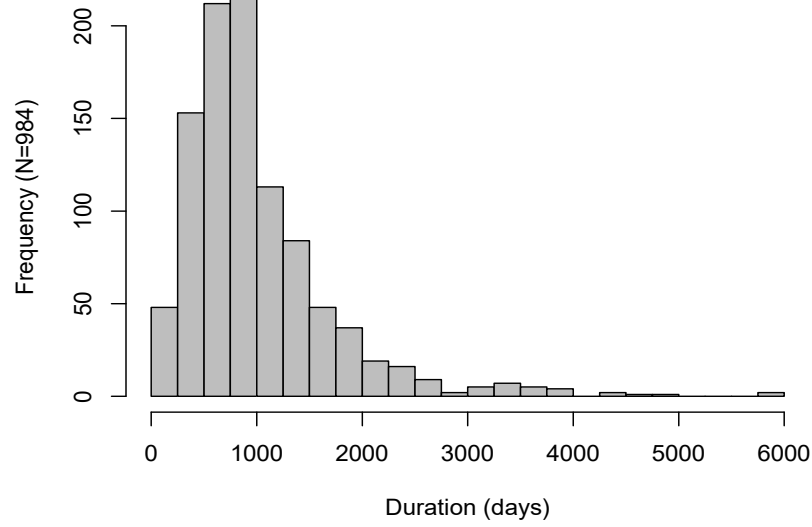


Figure 1: Histogram of Case Duration

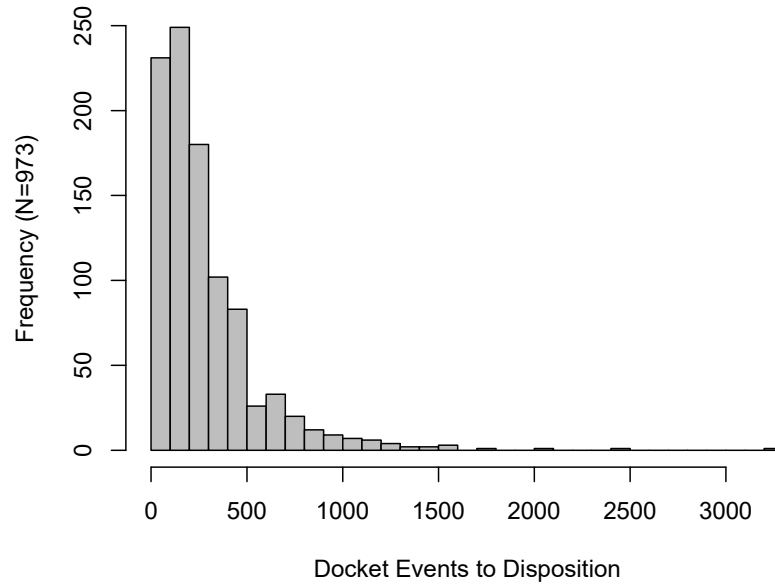


Figure 2: Histogram of Docket Events to Disposition

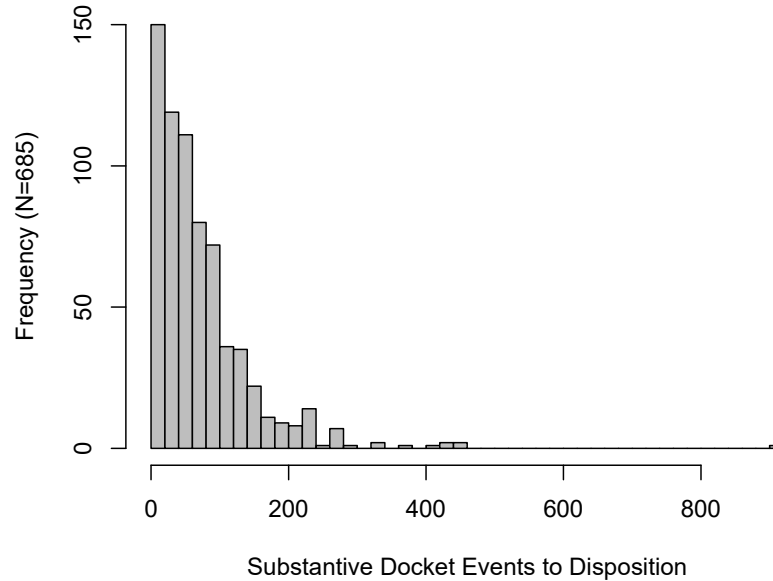


Figure 3: Histogram of Substantive Docket Events

C. Analysis of Complexity Metrics

First, we analyze overall case complexity based on outcome and disposition to determine whether patent holder wins are more or less complex than accused infringer wins. Next, we focus on the relative efforts of each party and further parse these results based on specific disposition on the asserted patent (*i.e.*, infringed, invalid, unenforceable or non-infringed).

1. Case Complexity by Outcome and Disposition

Figures 4 and 5 below show the breakdown of our metrics by case outcome. Figure 4 shows the total case duration, total number of docket entries, and number of substantive docket entries in patent holder wins versus accused infringer wins.

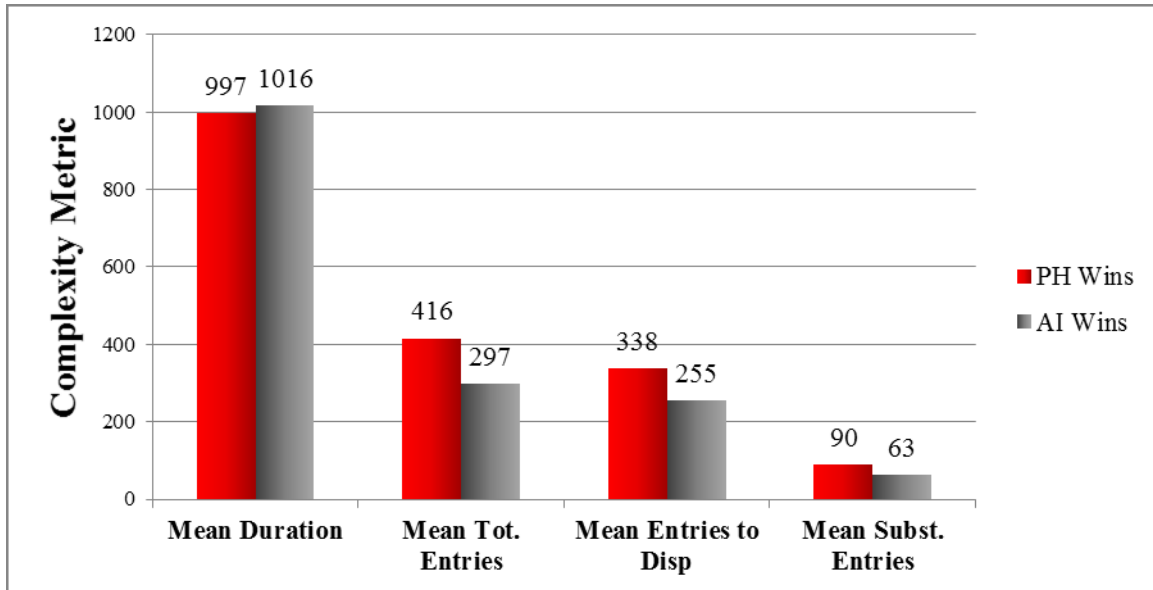


Figure 4: Case Complexity by Outcome (Overall Complexity)

One initial observation from Figure 4 above is that each of the complexity metrics based on docket entries is substantially greater for patent holder wins than in cases which the accused infringer prevailed.²⁵ Significance testing reveals that each of these differences is statistically significant at the 5% level.²⁶ This makes sense, given that patent holder wins involve an additional remedies phase, whereas damages and injunctions are typically not assessed in patent cases until infringement has been established. Determination of damages is often particularly complex and contentious, often requiring additional briefing, expert witnesses, specific findings of fact, and court rulings.

Turning from outcome to case disposition, we measure the complexity of cases finding infringement (patent holder wins) relative to invalidity, unenforceability and non-

²⁵ Notably, the fact that case duration does not follow suit suggests that, in line with our theory, duration is not a reliable metric of true complexity.

²⁶ The results of t-tests applied to log-transformed data of each metric are as follows:

Metric	Means	p-value
<i>Case Duration</i>	PH wins: 977d AI wins: 1016d	0.79
<i>Total Docket Entries</i>	PH wins: 416mo AI wins: 297mo	3.18e-9 ***
<i>Docket Entries to Disp.</i>	PH wins: 338mo AI wins: 255mo	1.73e-7 ***
<i>Substantive Docket Entries</i>	PH wins: 90mo AI wins: 63mo	3.02e-6 ***

infringement. Figure 5 illustrates this data. Whereas patent holder wins are generally more complex than each category of accused infringer wins, we observe that cases finding unenforceability tend to have higher numbers of docket entries and motions/orders. However, given that unenforceability is found for fewer than 10% of patents,²⁷ the increased average complexity likely reflects the particular features of these specific few cases relative to the larger dataset.²⁸

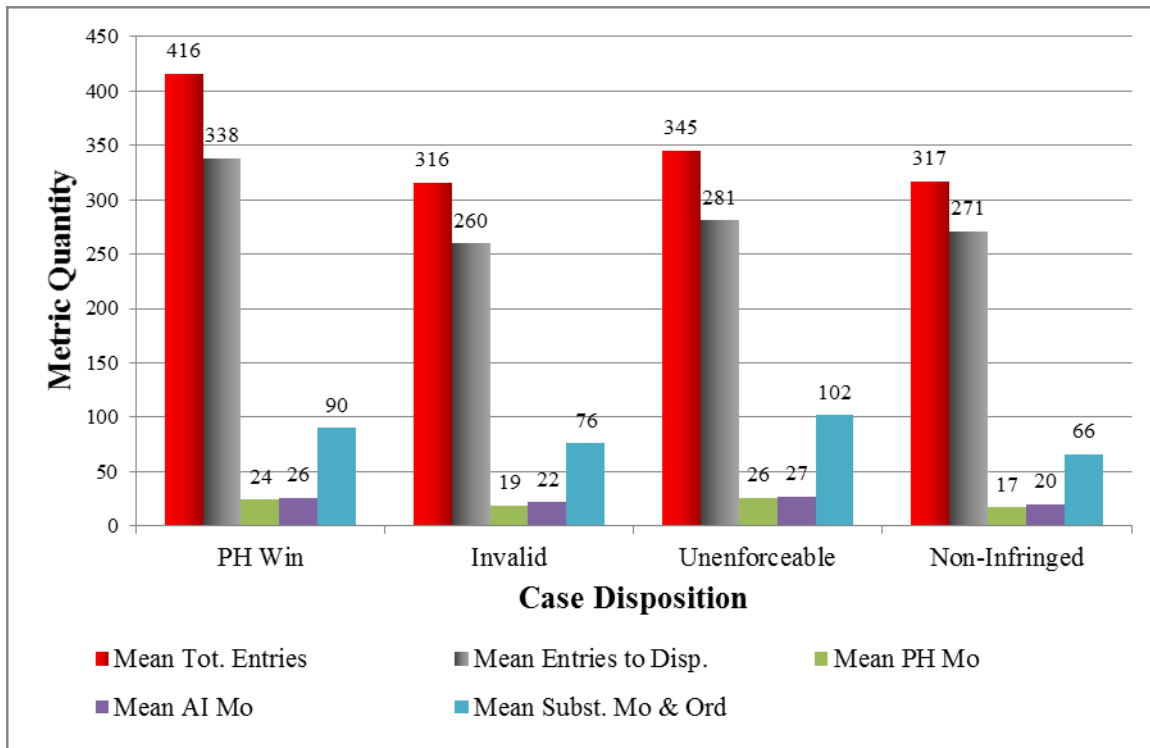


Figure 5: Case Complexity by Disposition (Each Metric)

2. Patent Holder vs. Accused Infringer Motions

Next we measure relative party effort in terms of the number of substantive motions filed by patent holders compared with accused infringers. Figure 6 shows a scatter plot mapping accused infringer motions against patent holder motions in each case. As this graph shows, the parties' respective motions appear to be strongly correlated with each other, with the number of accused infringer motions increasing with the number of patent holder motions (and vice versa). Figure 6 shows an overlay of the $y=x$ line representing equal numbers of motions by each party. The data generally lies

²⁷ See Appendix A, *infra*.

²⁸ We omit case duration from the results presented below, but we confirmed that case duration did not provide reliable measures in any of these analyses.

slightly above this line and appears to follow the same slope—this corroborates the observed correlation. However, the fact that most cases appear to lie above the $y=x$ line also suggests that the number of accused infringer motions may be slightly higher on average than the number of patent holder motions.

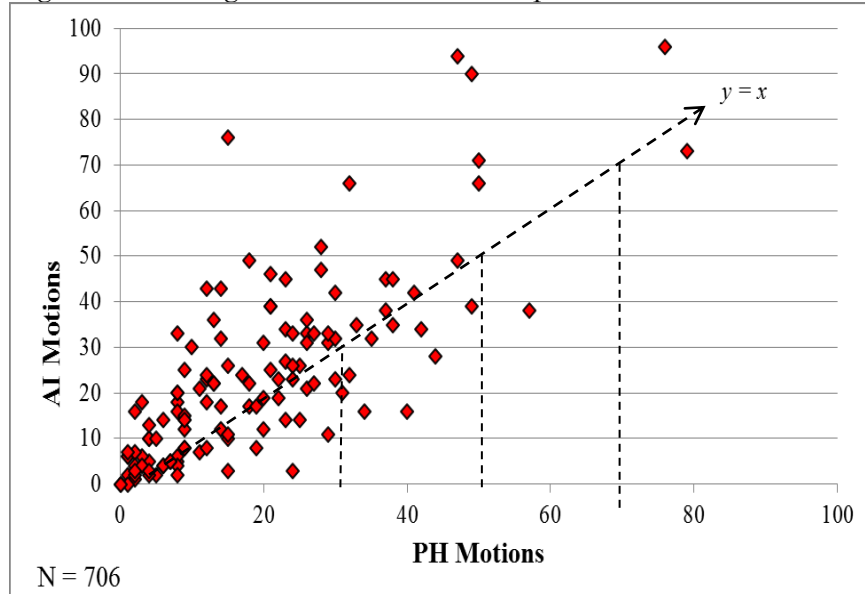


Figure 6: Patent Holder Motions vs. Accused Infringer Motions (Scatter Plot)

Figure 7 further illustrates this relationship, splitting patent holder wins from patent holder losses. In each grouping, accused infringers file more motions on average than patent holders. Yet the differences between the averages is relatively small, with approximately two additional motions filed on average by accused infringers in both sets of cases.

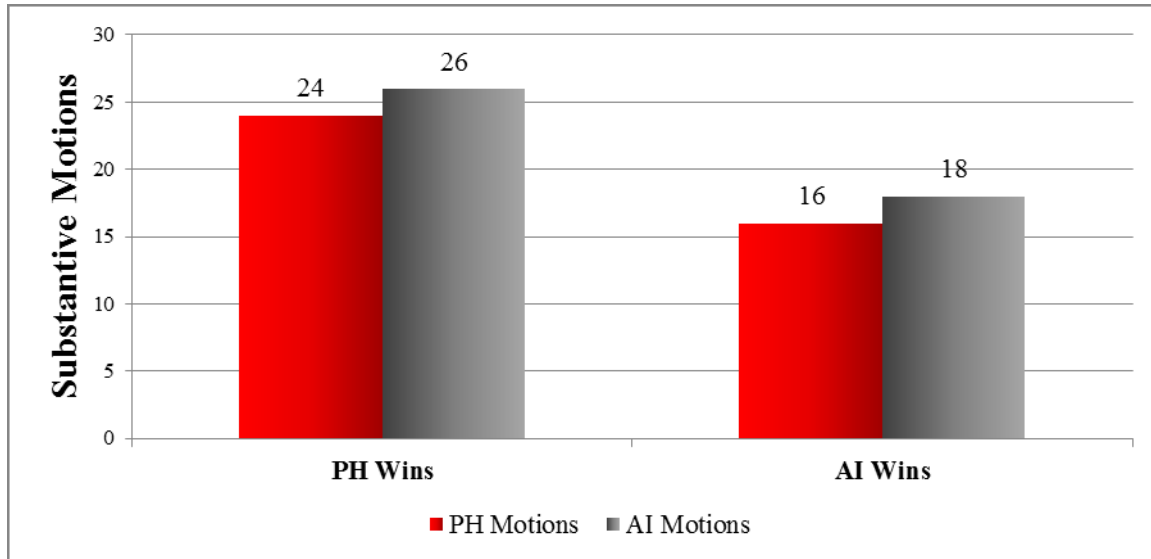


Figure 7: Patent Holder Motions vs. Accused Infringer Motions by Outcome

Significance testing reveals that the difference in each set of cases is significant only at the 10% level. Furthermore, significance is lost if a single motion is removed from the accused infringer cases in the aggregate. We expect that this corresponds to an additional motion for summary judgment filed by accused infringers. That is, given that the infringement analysis required for the patent holder to prevail is highly fact intensive, it typically requires adjudication at trial. In contrast, patent validity may be restricted to questions of law and may therefore succeed at summary judgment. Accordingly, the data suggests that accused infringers are likely availing themselves of summary judgment to try to invalidate the patents asserted against them.²⁹

This analysis also reveals that the determination of which party prevails in a case is not principally governed by which party filed more motions. We see that, as discussed *supra*, both patent holder wins and accused infringer wins involve relative parity in the average number of motions filed by each party. This is an encouraging result, as it suggests that the merits of the case, rather than simply the number of motions filed by the prevailing party, drive case outcome. By contrast, if we observed that the prevailing party is predominantly the party filing the most motions, that would suggest that litigation effort is the principal determinant of outcome, irrespective of the facts at hand. Rather, our observation that average party effort is equivalent between cases in which patent holders win and those in which accused

²⁹ *Accord* Kesan and Ball, *supra* note 5, at 264.

infringers prevail preserves the possibility that substantive merits of the case are important to the eventual decision.

IV. EMPIRICAL ANALYSIS

A. Complexity by Phase of Litigation

In this Section, we analyze the breakdown of case complexity according to the principal phases of patent litigation. We define these phases as follows: Phase I, complaint to claim construction (*Markman* hearing); Phase II, claim construction to disposition (summary judgment or trial); and Phase III, post-disposition proceedings (e.g., post-trial motions and orders). According to this convention, Phase I accounts for discovery and claim construction, as well as general case administration. Phase II comprises pre-trial scheduling and conferences (where applicable), briefing on motions for summary judgment (where applicable), interlocutory appeals from the *Markman* hearing, and preparation for the trial or summary judgment proceedings. Phase III includes motions for costs and fee-shifting as well as motions for Judgment as a Matter of Law (JMOL), new trial and remittitur, etc.³⁰

In Table 2, we report summary statistics of the number of docket entries in each phase of litigation. Phase II, *Markman*-to-disposition, involves the greatest number docket events, but Phase I, discovery and pre-*Markman* proceedings, are nearly as complex:

<i>Phase</i>	<i>Q1</i>	<i>Med.</i>	<i>Mean</i>	<i>Q3</i>
Pre- <i>Markman</i> (I)	69	104	155	183
Disposition (II)	65	149	195	249
Post-Disposition (III)	12	28	65	77

Table 2 – Docket Counts of Litigation Phases

³⁰ We identify each phase by first selecting all cases from our dataset which involve a separate *Markman* hearing (which, as shown in the regression results in Part C *infra*, are among the most complex cases). We then code the specific position of the *Markman* hearing in the case docket as a percentage of the total docket entries from the original complaint until disposition (summary judgment or trial). Finally, in order to capture the post-disposition phase, we calculate the difference in docket entries from the disposition to case closure phase. See Stuart J.H. Graham & Nicolas Van Zeebroeck, *Comparing Patent Litigation Across Europe: A First Look*, 17 STAN. TECH. L. REV. 655, 663 (2014) (describing the patent litigation process generally). Many thanks to the author for sharing this resource.

Next, we investigate the relative complexity of Phases I and II in patent holder wins vs. losses. Figure 8 below shows this analysis by mapping the position of the *Markman* hearing as a percentage of the number of docket entries between the complaint and disposition phases. We have ordered the cases according to increasing Phase I complexity and have differentiated patent holder wins (red squares) from losses (blue circles), with the average *Markman* position shown in black dashes.

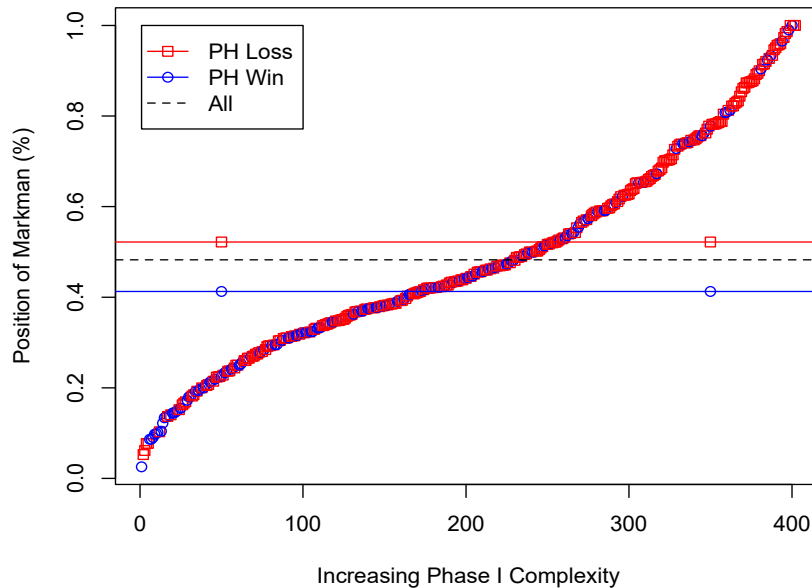


Figure 8: Phase I-II Complexity (PH Wins vs. Losses)

The difference in the average *Markman* position between patent holder wins and losses is statistically significant,³¹ indicating that the *Markman* hearing occurs proportionately earlier in patent holder wins than in losses. This corroborates the findings above regarding the increased complexity of patent holder wins corresponding to the remedies phase at trial—this results in a relatively longer Phase II and earlier *Markman* position in the docket as a whole.

Next, we analyze the relative complexity of Phase III, again distinguishing patent holder wins from losses. Figure 9 illustrates the data. Patent holder wins exhibit increased complexity in post-trial proceedings,³² which again may relate to ongoing litigation over remedies (e.g., post-trial motions to vacate or remit a jury award of damages). This also likely reflects the fact that patent holder wins generally require a trial, which can involve extensive post-trial proceedings; patent

³¹ p-Value of 1.90e-6.

³² The difference is significant at the 1% level (p-value of 0.006).

holder losses, by contrast, may be adjudicated at summary judgment and entail less ongoing litigation following the summary judgment order.

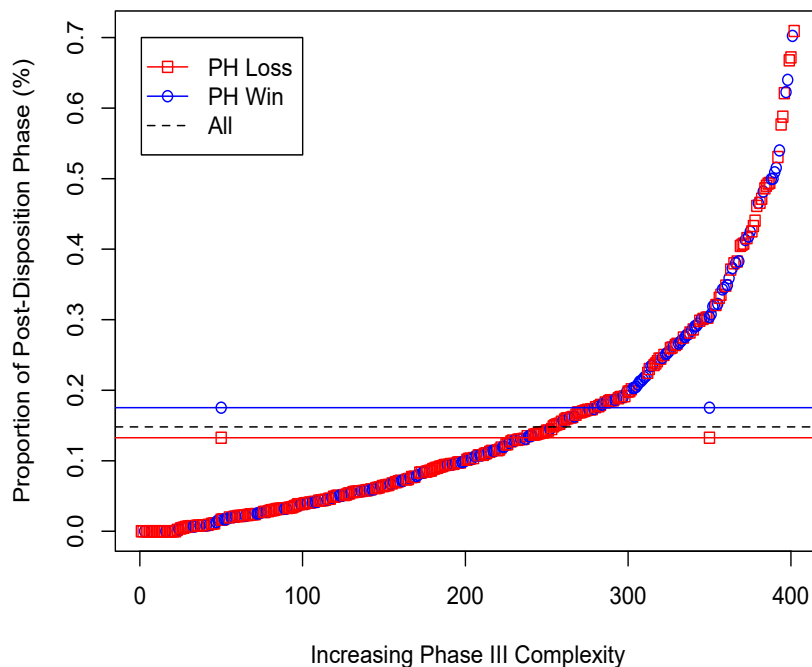


Figure 9: Phase II-III Complexity (PH Wins vs. Losses)

Finally, we analyze all three phases together to illustrate the relative complexity of each one. Figure 10 below shows the resulting graphs. We distinguish patent holder wins from losses by separating the cases into two sets of figures, and in each set we successively order the cases by increasing complexity in each of the three phases (for a total of six graphs). The vertical dashed line in these figures indicates the median case according to each ordering, and the final row provides box-plot comparisons of the data in each set.³³

By comparing the median positions in patent holder wins relative to losses, we see that Phase I (discovery and *Markman*) in wins is less complex than in losses, whereas Phases II (trial or summary judgment) and III (post-disposition) are each more complex in wins than losses. This demonstrates again that the differential complexity of patent holder wins tends to reside in the trial and post-trial phases, whereas patent holder losses tend to concentrate in relatively more complexity in Phase I and are resolved more summarily thereafter.

³³ For example, the line in the top-right figure represents the patent holder win having the median Phase III complexity.

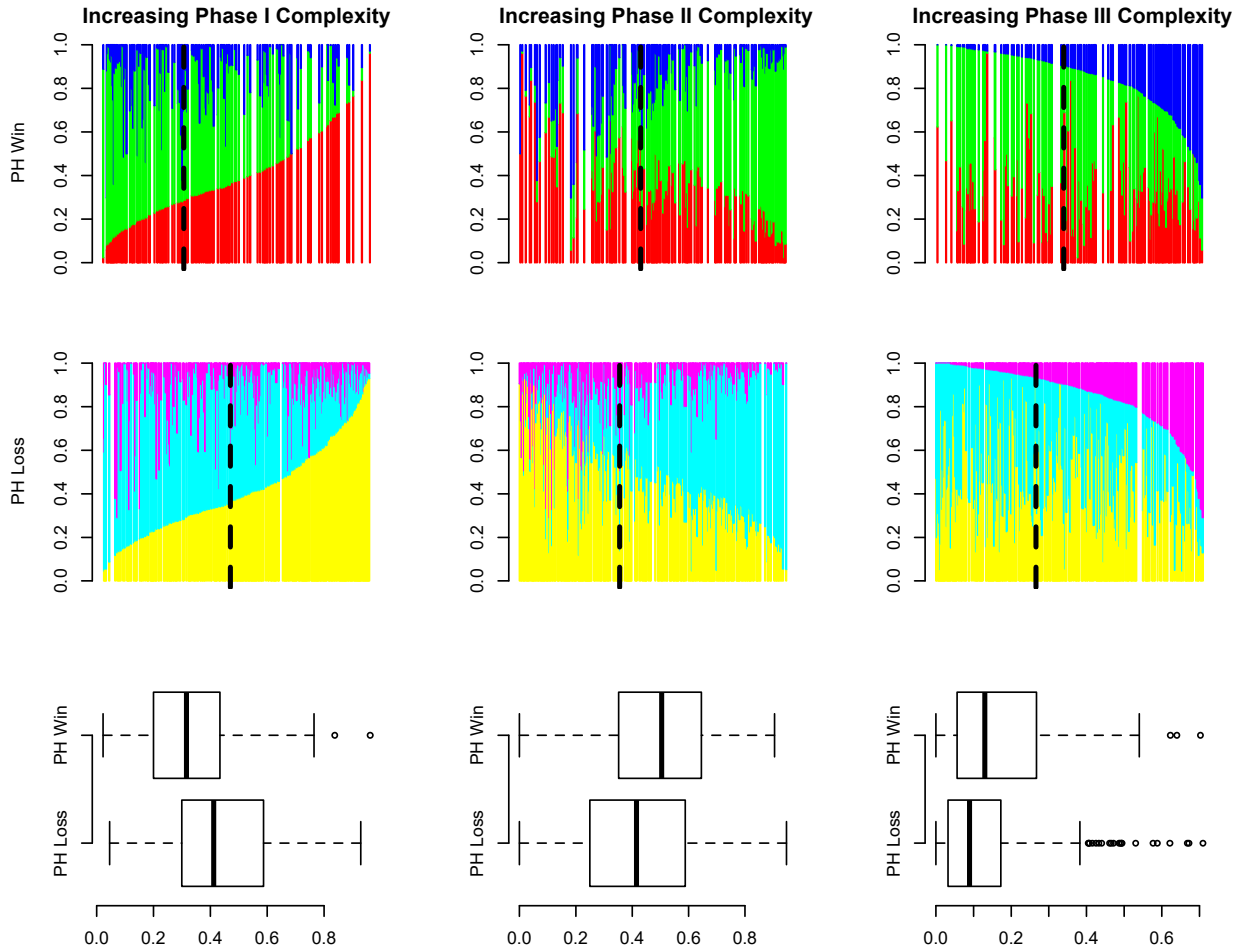


Figure 10: Complexity by Phase (PH Wins v. Losses)

B. Regression Analysis

Next, we conduct regression analyses to determine whether patent litigation complexity can be modeled statistically and, if so, which factors are principally correlated with overall case complexity.

1. Regression Models

We construct linear regression models starting with a baseline set of standard variables and iteratively add new factors and test whether they increase the explanatory power of the resulting model. Table 3 below shows the overall fit parameters of the final models for each metric:

<i>Dependent Variable</i>	<i>R²</i>	<i>Adj. R²</i>	<i>Std. Error</i> <i>(df)</i>	<i>F Statistic</i> <i>(df1, df2)</i>	<i>p-value</i>
(1) Docket Entries to Disposition	0.460	0.427	0.697 (911)	14.09 (55, 911)	< 2.2e-16
(2) Total Docket Entries	0.477	0.445	0.674 (911)	15.09 (55, 911)	< 2.2e-16
(3) Subst. Docket Entries to Disposition	0.455	0.403	0.686 (576)	8.74 (55, 576)	< 2.2e-16
(4) Duration to Disposition	0.196	0.149	0.616 (921)	4.09 (55, 921)	< 2.2e-16

Table 3 – Regression Models of the Complexity Metrics

Appendices C through F provide the full regression results of certain relevant variables corresponding to the models above. One immediate observation is that each docket-based metric has considerably better fit than time duration.³⁴ This is consistent with the theory that duration is a noisy metric for litigation, and particularly for patent cases that may be subject to stays and other breaks in continuity.³⁵

Each of the docket-based metrics exhibits a reasonably good fit, which suggests that our models are robust and largely complete.³⁶ However, litigation dockets exhibit a high degree of idiosyncratic variation case-to-case, which limits the best degree of fit achievable by any statistical model.³⁷

³⁴ Specifically, although all of the p-values are statistically significant at standard levels, the duration p-value is larger (less significant) and the corresponding F statistic is three to four times lower than those of the docket metric models. The values for multiple R² and adjusted R² are also substantially greater for each of the docket metrics.

³⁵ See Kesan and Ball, *supra* note 5, at 281 (“Time to termination is a traditional measure of the resources expended on a court case. However, while it has a strong intuitive appeal, this measure is also likely to be inaccurate. There can be long delays in scheduling court hearings and periods of inactivity that are not necessarily associated with higher costs. The number of documents filed in the case is probably more closely correlated with actual costs, particularly in the form of ‘billable hours’ of attorney time.”).

³⁶ In particular, each has a relatively high F statistic (F values increase from 1) and correspondingly near-zero p-value. The residual errors of each model appear to be normally distributed and do not exhibit clear, non-random trends, as shown in the graphs in Appendices C-E, which suggests that there are no strong determinative factors missing from the models. Finally, the multiple and adjusted R² values of each model, typical measures for the degree of fit, are reasonably high.

³⁷ Other features of the models also suggest that we are capturing the bulk of the non-idiosyncratic factors that affect complexity in each case and are not missing key variables. For example, a consistent set of factors are significant across the models for all metrics; furthermore, the results of each model are generally tolerant to minor changes in the selection of variables. Additionally, the errors of each model are generally normally distributed.

2. Significant Factors Associated with Case Complexity

Next, we analyze how certain key factors relate to case complexity. We focus here only on the total docket count models, given their greater degree of fit than the models of case duration and substantive docket entries.

(a) Case Procedure and Disposition Variables:

The case procedure variables generally behave consistently with expectations. For example, cases that were stayed and then resumed, such as stays for parallel USPTO or ITC proceedings involving the asserted patents, tend to involve more docket filings than uninterrupted cases. This likely reflects the heightened contentiousness of the dispute—the parties are litigating across multiple forums—as well as the possibility of changed circumstances when the case resumes and must be briefed and analyzed anew.

Also, cases that involve separate *Markman* hearings for claim construction involve more docket activity, which reflects additional litigation effort for such proceedings. In turn, cases involving motions for interlocutory appeal by either party following a *Markman* hearing also reflect greater complexity.

Trials are statistically significant, and we also observe greater complexity of cases involving jury trials relative to bench trials. This makes sense—jury trials require *voir dire* proceedings, and may also involve disputes over jury instructions and other trial procedures. Also, it is commonly understood that bench trials are typically not opted for in the most high-stakes patent disputes, and therefore the presence of jury trials likely coincides with other factors contributing to case complexity.

Finally, we also observe some evidence of Circuit variations, with U.S. District Courts located in the Third Circuit and Ninth Circuit exhibiting greater complexity in their cases on average. The significance of the Ninth Circuit likely reflects a greater proportion of high-tech and software patent cases in the Northern and Central Districts of California near Silicon Valley, whose cases involve complex issues of validity and changing standards governing the patentability of computer software. These cases may also exhibit greater overlap with trade secret, copyright and other issues in the high-tech fields.

(b) Party Characteristics:

The most interesting set of results involve the types and characteristics of parties who are involved in cases exhibiting the highest complexity. For example, the size of the accused infringer registers as strongly and positively associated with increased case complexity. This may indicate that large corporate defendants tend to fight harder, and the process of enforcing patents against such entities may be more difficult. Also, litigation involving large defendants likely entails greater evidentiary complexity, including more document production, corresponding motions in limine, and court rulings on exclusion, privilege, and other issues.

Similarly, the size of the patent holder is significant and positively associated with increased complexity.³⁸ This could indicate that large corporate patent holders prosecute their claims more aggressively, but it also likely reflects similar size with respect to discovery, determination of damages and adjudication of other issues.

One result that reflects recent policy focus is the increase in case complexity associated with the number of accused infringers. This likely reflects practices during the height of the patent litigation boom, whereby patent holders could sue an entire industry with a single complaint. The joinder rules of the AIA were designed to prevent such practices by requiring a greater nexus between independent defendants in order to join them in the same case.³⁹ Yet, these rules have also been criticized for increasing the overall number of lawsuits and potentially inhibiting aggregated disposition of common issues, which arguably may increase the overall burdens on the litigation system.⁴⁰ In any event, we observe that cases involving multiple accused infringers exhibit greater complexity, and post-AIA we expect that the propensity of such cases has been reduced.

Also important from a policy perspective is the complexity of litigation involving Non-Practicing Entities (NPEs). We separated NPEs who are individuals, universities, and companies (including Patent Assertion Entities (PAEs)). Our results reveal that cases involving NPE-individuals have significantly lower complexity on average. This may reflect the lesser resources available to individual inventors and the difficulties they face in protecting their rights against infringers. By contrast, we observe no significant difference in

³⁸ Additional research is being conducted to investigate whether these results can be further parsed by pairwise groupings of the parties, such as cases involving large companies on both sides.

³⁹ 35 U.S.C. 299 (2012).

⁴⁰ See, e.g., Dongbiao Shen, *Misjoinder or Mishap? The Consequences of the AIA Joinder Provision*, 29 Berkeley Tech. L.J. (2014).

the complexity of PAE litigation relative to other cases, based on these metrics.

(c) Patent Attributes:

Several interesting patent attributes relate to case complexity. For example, although the individual technology classifications of the asserted patents are not highly significant, we do observe that high-tech cases specifically are more complex than the average.⁴¹

Interestingly, forward citations, which are typically associated with patent value or importance, exhibit no significance. Forward citations of a patent are the aggregate number of future patents that cite to the original patent in their lists of “references cited”; several databases keep track of forward citations, including the USPTO’s “Referenced By” feature,⁴² and they are often used in statistical analyses as proxies for patent value, among other things. This is striking because one might expect cases involving higher value patents to be more contentious. However, it is possible that two effects are counteracting each other in the overall data—certain high value patents are less complex to litigate, as they are clearer and less susceptible to lengthy disputes over claim construction, whereas high value patents may generally involve increased case complexity due to higher stakes.

Patent family size is also strongly positively significant—this likely corresponds to higher-stakes litigation involving more developed technologies that have been improved and refined by the patent holder. Notably, to some extent, this may correlate with entity size as well, reflecting patent holders with larger R&D budgets.

Finally, one particular patent variable that we would expect to correspond to increased complexity is the number of patents involved in the case. When multiple patents are asserted, regardless of outcome, we would expect greater complexity than cases involving a single patent. Although our flag for multiple patents does not register as significant in the overall regressions, this may be the result of interactions with other variables—in particular we suspect that cases brought by and against large companies are likely to involve multiple patents, making it difficult to parse out the individual significance of each factor.

We test this hypothesis by creating interaction terms to identify cases where a large patent holder or accused infringer

⁴¹ The high-tech flag is based on technology categories assigned by PwC in the underlying dataset.

⁴² See www.uspto.gov.

was involved and whether single or multiple patents were asserted.⁴³ As shown below in Table 4, cases involving multiple patents are significantly more complex than cases involving a single patent:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
Multi-Patents x Large-Entity PH	0.24	0.08	2.99	0.002	**
Single-Patent x Large-Entity PH	0.03	0.08	0.40	0.68	
Multi-Patents x Large-Entity AI	0.14	0.07	1.92	0.06	.
Single-Patent x Large-Entity AI	0.11	0.07	1.63	0.10	
				R ² :	0.476
				Adj. R ² :	0.446
				Std.Err.:	0.674
				F _(52,914) :	15.95
				p-val:	2.2e-16
				N:	967

Table 4 –Regressions of Multi-Patent Interactions

⁴³ Results are consistent for both docket entry metrics; in the table below we report results using the total docket count dependent variable.

*Time Trend Analysis and Event Study***3. Time Trend Analysis**

Next, we investigate whether case complexity has significantly increased or decreased over time, across the years, in our sample. Table 5 below shows the results of a simple linear regression using our metrics as the dependent variable and the year of decision as the independent variable. We find a significant increasing time trend year-to-year in the number of docket entries (both total and number to disposition). Also, with respect to substantive docket entries, although we detect no significant time trend overall (2004-2009), we do find a slightly significant trend in the most recent years (2007-2009). By contrast, case duration does not exhibit a significant time trend. Figures 11 and 12 depict these trends via year-to-year box-plots.

<i>Metric</i>	<i>R²</i>	<i>Adj. R²</i>	<i>Std. Error (df)</i>	<i>F Statistic (df1, df2)</i>	<i>p-value</i>
Case Duration	5.36e-6	-1.01e-3	0.666 (981)	5.26e-3 (1, 981)	0.942
Tot. Docket Entries	0.0111	0.0101	0.915 (971)	10.88 (1, 971)	1.01e-3 **
Docket Entries to Disposition	0.00871	0.00768	0.900 (971)	8.56 (1, 971)	3.58e-3 **
Subst. Docket Entries	7.31e-5	-1.51e-3	0.888 (631)	0.0461 (1, 631)	0.830
Subst. Docket Entries ('07-'09)	7.43e-3	4.94e-3	0.876 (399)	2.986 (1, 399)	0.0847 .

Table 5: Time Trend Regressions of Complexity Metrics

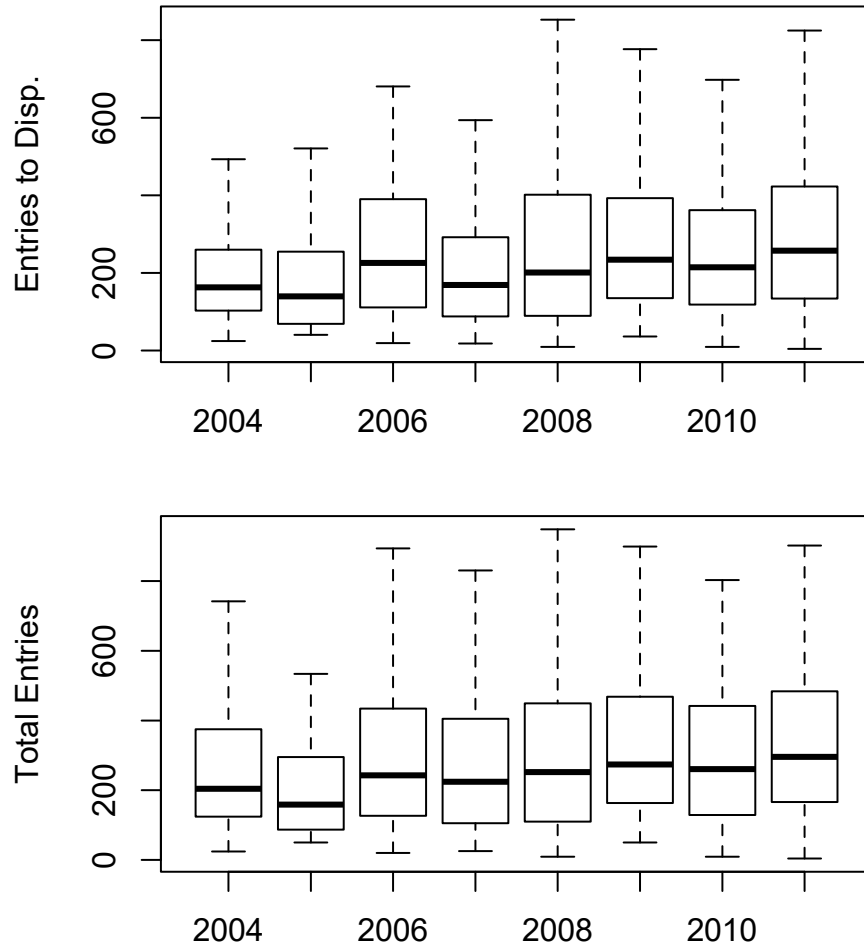


Figure 11: Box-Plots for Number of Docket Entries by Year of Disposition

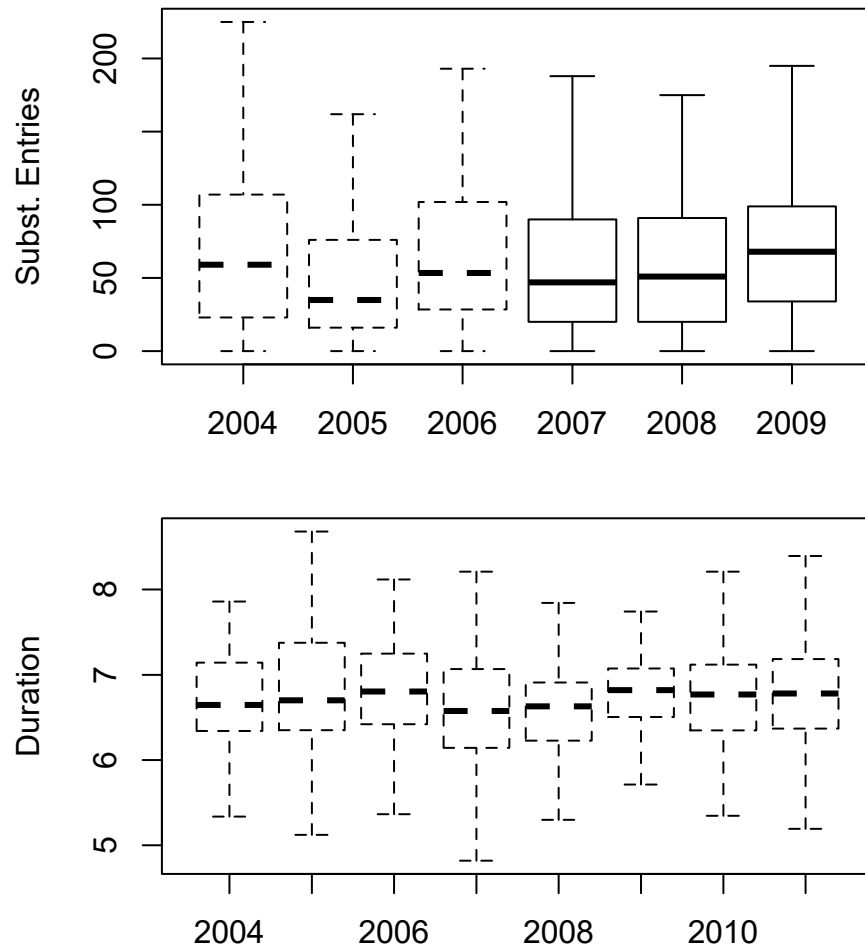


Figure 12: Substantive Docket Entries and Case Duration by Year of Disposition

The fact that we observe significant increasing time trends, particularly in recent years, is somewhat concerning. To some extent, this appears to reflect common concerns about the increasing burdens of patent litigation on defendants. We find corroboration by separating patent holder wins from losses, whereby only patent holder losses exhibit a significant time trend. Figure 13 and Table 6 show this result:

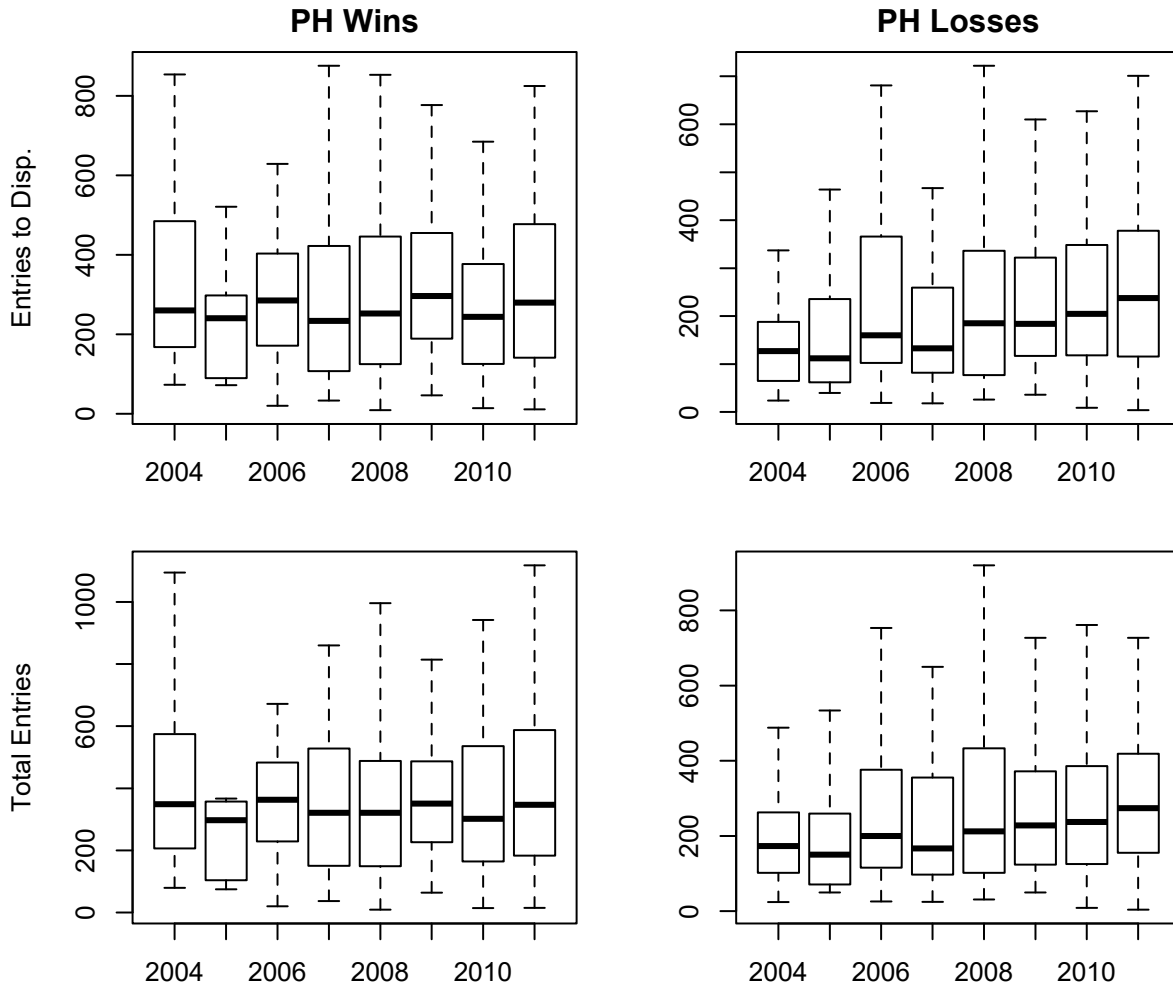


Figure 13: Time Trends by Prevailing Party

<i>Metric</i>	<i>R²</i>	<i>Adj. R²</i>	<i>Std. Error (df)</i>	<i>F Statistic (df1, df2)</i>	<i>p-value</i>
Entries to Disp. [PH Win]	1.76e-4	-2.85e-3	0.945 (331)	0.0581 (1, 331)	0.810
Entries to Disp. [PH Loss]	0.0254	0.0239	0.865 (638)	16.63 (1, 638)	5.12e-5 ***
Entries to Disp. [PH Win]	1.27e-4	-2.89e-3	0.945 (331)	0.0421 (1, 331)	0.838
Total Entries [PH Loss]	0.0188	0.0173	0.848 (638)	12.74 (1, 638)	5.00e-4 ***

Table 6: Time Trends by Prevailing Party

However, the fact that complexity of patent holder wins does not exhibit a significant time trend does not necessarily mean that the trend in patent holder losses reflects disproportionately increasing burdens on accused infringers. First, there are substantially fewer patent holder wins, which makes trends more susceptible to distortion by outliers and less

likely to register as significant when small differences are involved. Furthermore, since patent holder wins are more complex overall, the absence of a clear trend could indicate that certain policy or other changes over time have selectively reduced the complexity of patent holder wins. Particularly if complexity has generally increased across all cases (*e.g.*, corresponding with inflation), a selective complexity-reducing impact due to policy changes could register as the absence of a net time trend for patent holder wins. We test this hypothesis in Part 2 below.

4. Effect of Policy Changes

In the final analyses, we investigate whether major recent policy shifts have had significant effects on the complexity of patent cases. To do so, we conduct difference-in-difference (DID) regressions to determine whether the complexity of a specific set of cases has changed relative to another set of cases, controlling for all other factors. More particularly, the DID specification tests whether there is a significant difference before versus after a certain event (the first “difference” in “difference-in-difference”) between the extent to which the complexity of two distinct sets of cases is different from each other (the second “difference”). The event in question is expected to impact only one set of cases, and therefore the DID analysis is able to measure the extent of this impact and support an inference that the event caused it.

The two different sets of cases at play are patent holder wins and losses, respectively. The “event” in question is actually a series of policy shifts occurring during the 2008+ (2008-2011) timeframe, which principally took place via several Federal Circuit decisions regarding remedies for patent infringement. To the extent these decisions affected case complexity, we would expect to see a differential effect on patent holder wins, which are subject to these new standards, relative to patent holder losses, in which remedies are not adjudicated. Accordingly, we seek to compare the difference in complexity between patent holder wins and losses across two timeframes—pre-2008 (2004-2007) and 2008+ (2008-2011).

In particular, the Federal Circuit’s decisions in *Lucent v. Gateway*,⁴⁴ *ResQNet v. Lansa*,⁴⁵ *Cornell v. HP*⁴⁶ and *Uniloc v. Microsoft*,⁴⁷ among other cases, heightened the standards for proving lost profits and reasonable royalty damages from

⁴⁴ *Lucent Techs., Inc. v. Gateway, Inc.*, 580 F.3d 1301, 1301 (Fed. Cir. 2009).

⁴⁵ *ResQNet.com, Inc. v. Lansa, Inc.*, 594 F.3d 860 (Fed. Cir. 2010).

⁴⁶ *Cornell Univ. v. Hewlett-Packard Co.*, 609 F. Supp. 2d 279 (N.D.N.Y. 2009).

⁴⁷ *Uniloc USA, Inc. v. Microsoft Corp.*, 632 F. 3d 1292, 1335 (Fed. Cir. 2011).

infringement and therefore may have increased the complexity of recovering remedies in cases where liability is established.⁴⁸ For example, in *Lucent v. Gateway*, the Federal Circuit vacated a jury's reasonable royalty award of a lump-sum amount based on several prior license agreements, requiring a more careful analysis of the applicability of the licenses to the context at hand and scrutinizing the sufficiency of the evidence supporting the jury's award.⁴⁹ In *ResQNet*, the Federal Circuit again emphasized that proof of damages "requires sound economic proof of the nature of the market and likely outcomes with infringement factored out of the economic picture."⁵⁰ In *Cornell*, Federal Circuit Judge Rader, sitting by designation in a New York District Court, required "credible economic indicators" to prove lost profit damages, offering remittitur of the jury's damages award based on the "entire market value rule."⁵¹ Finally, in *Uniloc*, the Federal Circuit rejected the long-standing "25% Rule of Thumb," which was used to set a baseline royalty rate as a starting point of damages calculations.⁵² Together, these cases arguably represent a heightening of the burdens of proving patent infringement damages, potentially increasing the litigation effort required to recover. Accordingly, cases that awarded damages following these decisions should exhibit increased complexity, and that complexity should be specifically situated in the remedies phase. We test these hypotheses below.

We construct the DID model with certain Boolean flags that identify (1) patent holder wins versus losses, (2) cases decided pre-2008 and 2008+, and (3) patent holder wins specifically in the 2008+ timeframe (an interaction variable of the previous two flags). We are interested in the sign and significance of the coefficient of the third variable—a significant positive coefficient will mean that the complexity of patent holder wins relative to losses (*i.e.*, the overall complexity of cases in which patent holders win relative to the complexity of cases in which accused infringers win) has increased across the two time periods, and a significant negative coefficient will indicate a decrease.

Based on the results in preceding Part, we expect to see a *decrease* in the complexity of wins relative to losses from pre-2008 to 2008+, given that patent holder losses exhibit

⁴⁸ See also *IP Innovation LLC v. Red Hat Inc.*, No. 2:07-CV-447 (RRdR), 2010 WL 986620 (E.D. Tex. Mar. 2, 2010); *WordTech Sys., Inc. v. Integrated Network Solutions, Inc.*, 609 F.3d 1308, 1319 (Fed. Cir. 2010).

⁴⁹ *Lucent*, 580 F.3d at 1329-30.

⁵⁰ *ResQNet*, 594 F.3d at 870 (quoting *Grain Processing Corp. v. Am. Maize- Prods. Co.*, 185 F.3d 1341, 1350 (Fed. Cir. 1999)).

⁵¹ *Cornell*, 609 F. Supp. 2d at 288.

⁵² *Uniloc*, , 632 F. 3d at 1335.

increasing complexity over time whereas patent holder wins have no overall time trend. Table 7 shows this result, based on a simple regression involving only flags (1)-(3) above:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
1) Patent Holder Win?	0.391	0.10	4.09	4.8e-5	***
2) Case Year 2008-2011?	0.234	0.07	3.15	1.7e-3	**
3) Patent Holder Win x 2008-2011?	-0.197	0.12	-1.59	0.112	

R²: 0.036
 Adj. R²: 0.032
 Std.Err.: 0.860
 F_(3,861): 10.61
 p-val: 7.4e-7
 N: 874

Table 7 –Difference-in-Difference Model (without Fixed Effects)

We see from Table 7 that the coefficient on the interaction term (3) is in fact negative, corresponding to the relative decrease in the complexity of wins during this period, but its magnitude is not statistically significant.⁵³ In absolute terms, the difference in number of docket events decreased from approx. 243 for wins vs. 164 for losses pre-2008 to approx. 252 for wins (a negligible increase) to 207 for losses (a significant increase) in the years 2008 onwards.

In order to ascertain whether the different trends of wins vs. losses is indeed attributable to the aforementioned policy changes in the 2008+ period, we must control for other factors that influence case complexity. Adding the time-invariant case, party and patent characteristics (fixed effects) from our final regression model above allows us to do this. Table 8 shows the resulting regression coefficients of the relevant Boolean flags in this model:

	<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
1)	Patent Holder Win?	0.0783	0.0870	0.900	0.368	
2)	Case Year 2008-2011?	0.109	0.100	1.091	0.275	
3)	Patent Holder Win x 2008-2011?	-0.307	0.101	-3.049	0.0024	**
4)	[Combined Fixed Effects]	4.099	-	-	-	-
					R ² :	0.435
					Adj. R ² :	0.407
					Std.Err.:	0.674
					F _(40,820) :	15.75
					p-val:	2.2e-16
					N:	874

Table 8 –Difference-in-Difference Model (with Fixed Effects)

Here, after controlling for other factors of the litigation we see that patent holder wins (relative to losses) in the 2008+ time period are significantly less complex than pre-2008, all else equal. We interpret this result as indicating that the policy shifts regarding patent remedies have reduced the complexity of patent holder wins.⁵⁴

5. Parsing the Difference

Finally, we investigate what phases of litigation produced the change in relative complexity of patent holder wins to losses that we observe above, using the phase delineations defined in our earlier analysis.⁵⁵ Figure 14 below shows the change in docket counts from pre-2008 to 2008+ cases across each phase of litigation, with patent holder wins in the top row and patent holder losses in the bottom row:

⁵⁴ It is possible to interpret this event another way, namely that patent holder losses have *increased* as a result of policy shifts or other exogenous events during the 2008-2011 timeframe. However, we are not aware of any such events. Rather, given the comprehensive set of variables coded and absence of other explanations, we think that the more likely explanation is that case complexity overall has increased with time, but recent policy events have simultaneously had a decreasing effect on the complexity of patent holder wins.

⁵⁵ This analysis is only conducted for the subset of cases involving a separate *Markman* hearing.

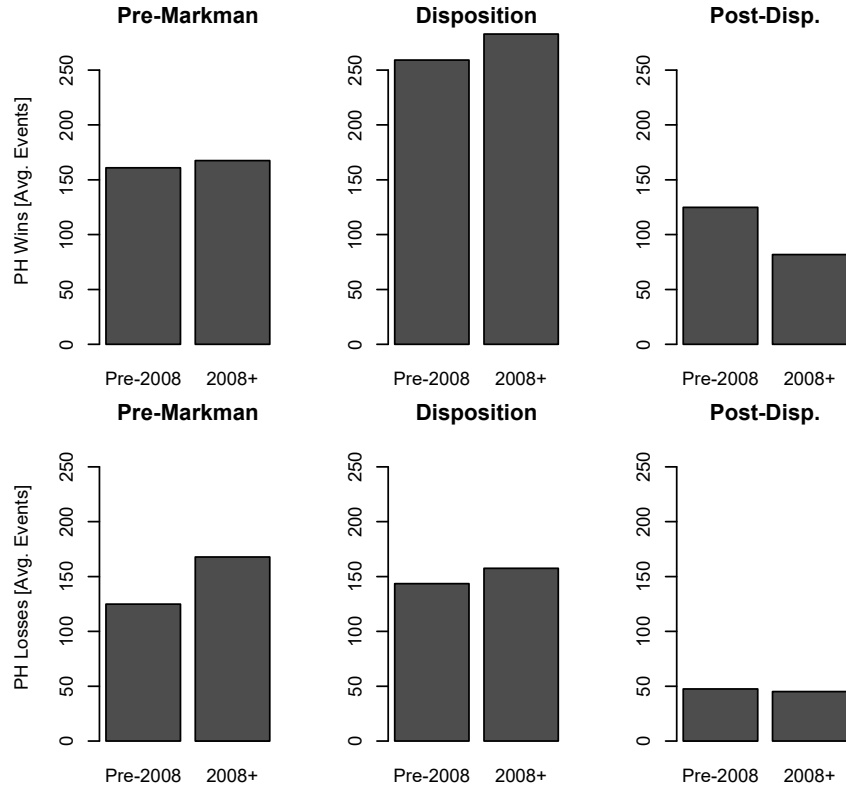


Figure 14: Changes in Complexity by Phase of Litigation (Wins vs. Losses Separated)

The results are striking. Regarding patent holder wins, we observe an apparent increase in complexity of Phase II (trial), as well as a decrease in the post-trial proceedings of Phase III.⁵⁶ This suggests that the policy shift in remedies case law starting around 2008 has made disputes over damages and injunctions more contentious. That is, having controlled for all other factors in the DID specification, we interpret the escalation in Phase II complexity as increased litigation effort regarding adjudication of remedies, namely damages and injunctions. Also, the decrease in Phase III could indicate that resulting remedies are less favorable to patent holders, giving rise to fewer disputes involving motions for JNOV, remittitur and other post-trial activity. Additionally, accused infringers may be moving directly to appeal, thinking that the Federal Circuit will reverse the District Court under the new remedies jurisprudence.

Additionally, we find that the increase in complexity of patent holder losses over time is driven principally by increasing Phase I complexity.⁵⁷ This could reflect heightened

⁵⁶ These differences are not statistically significant at the 5% level; however, the sample sizes are quite small for each subset.

⁵⁷ Significant at the 1% level (p-Value of 0.00577).

complexity of discovery proceedings, perhaps corresponding to proliferation of e-discovery and document retention practices. However, this also likely reflects increased effort by defendants to secure claim constructions that drive the final outcome in their favor.⁵⁸

Finally, we report the net change in the complexity of each phase across all cases (combining wins and losses). As shown in Figure 15, both Phases I and II exhibit increases in complexity.⁵⁹

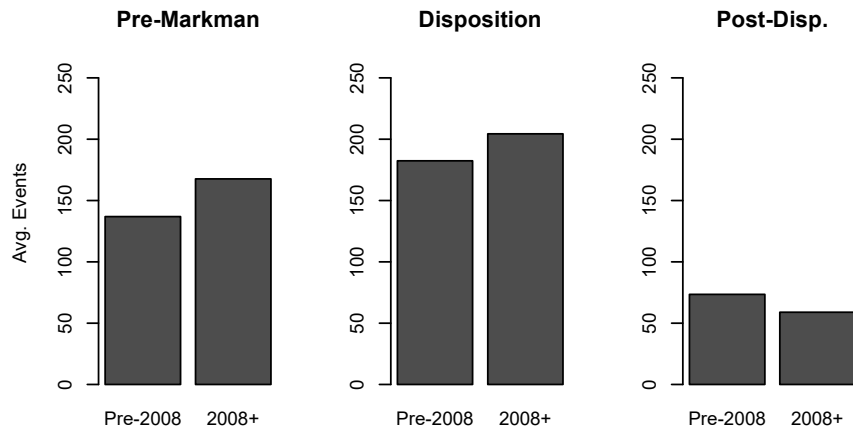


Figure 15: Changes in Complexity by Phase of Litigation (All Cases Combined)

V. CONCLUSION

In this study we investigate patent case complexity using five principal lines of inquiry. First, at the most basic level we ask how complexity can be measured, so as to support detailed statistical analysis. We find that docket activity yields the most robust metrics for analysis. We construct three metrics based on docket activities, reflecting the numbers of docket entries to disposition and case closure, respectively, as well as substantive motions by the parties and court orders.⁶⁰

Second, we ask how complexity varies across different types of cases. We find that patent holder wins are significantly more complex than patent holder losses. This added complexity corresponds to the remedies phase, which is generally present only in patent holder wins. Notably, individual party effort (in terms of number of motions filed)

⁵⁸ Although there appears to be a slight increase in Phase II complexity of patent holder losses, the change is not significant at the 5% level.

⁵⁹ However, only the increase in Phase I complexity is significant at the 5% level (p-Value = 0.0204), whereas the increase in the Phase II complexity is significant only at the 15% level (p-Value = 0.1147). The observed decrease in Phase III is not significant.

⁶⁰ As expected, our fourth metric, total case duration, does not accurately reflect case complexity.

does not significantly differ between patent holder wins and losses, which suggests that case outcome is not driven solely by which party litigates harder.

Third, we investigate the complexity of the three principal phases of patent litigation, as punctuated by the *Markman* claim construction hearing and the dispositive District Court order. We find that Phase II (trial or summary judgment) is the most complex, followed closely by Phase I (discovery and *Markman*). By contrast, Phase III (post-disposition) is significantly less complex. Also, we compare the respective phases between patent holder wins and losses, and we find that Phase II in patent holder wins are significantly more complex than in patent holder losses, reflecting remedies determinations in patent holder wins and the possibility of summary judgment dispositions in patent holder losses.

Fourth, we run regression analyses to identify the principal factors associated with greater or lesser complexity in each case. We are able to construct regression models with a reasonable degree of fit with respect to each of the docket entry metrics. Among the most notable results, party size has a significant increasing effect on case complexity, for both patent holders and accused infringers. Looking more closely at the complexity of NPE litigation, we find that PAE cases are not significantly more complex than other cases on average, whereas cases involving individual plaintiffs tend to be significantly less complex, possibly reflecting lower resources and sophistication of individuals. Certain patent attributes are also significant, including cases involving computer technology (high tech) patents and cases in which multiple patents are at issue.

Finally, we ask whether complexity has increased over time, and we further analyze the impact of key policy shifts in recent years. We find that complexity has increased over time in the aggregate, which appears to be driven by an increase in the complexity of discovery and *Markman* proceedings, particularly in cases where accused infringers prevail. This could reflect the advent of e-discovery proceedings; alternatively, it could also suggest that accused infringers are litigating more aggressively prior to claim construction.

Moreover, we find that key policy shifts by the Federal Circuit and Supreme Court to reform their jurisprudence of patent remedies have had a significant impact on the complexity of patent holder wins. Overall, the complexity of patent holder wins relative to losses has decreased from the 2004-2007 to 2008-2011 time periods, and the change is statistically significant after controlling for other relevant factors. By decomposing each set of cases into their litigation phases we can further investigate where the changes in complexity occurred, and we find that the complexity of trials

in patent holder wins has generally increased, whereas the complexity of post-trial proceedings in those cases has decreased.

We conclude that the determination of patent infringement remedies has become more contentious and complex as a result of the Supreme Court's and Federal Circuit's policy shifts. Cases such as *Uniloc* (invalidating the "25% rule of thumb" for calculating reasonable royalties),⁶¹ *Cornell v. HP* (tightening the requirements to prove lost profits damages),⁶² *ResQNet* (expanding evidentiary bases for challenging reasonable royalties),⁶³ and *Lucent v. Gateway* (interpreting various *Georgia-Pacific* factors for determining reasonable royalty rates),⁶⁴ have driven this trend. Importantly, trials remain the sole venue for patent holders to enforce and recover from infringement of their rights. To the extent patent trials have become too complex, or excessively skewed against patent holders, the value of patents and the innovation capital they provide could be harmed.

Looking forward, this study opens a number of avenues for future analysis. In particular, the dramatic recent changes to the U.S. patent litigation system under the AIA are likely to have significantly affected case complexity. Analyzing the complexity of recent cases once sufficient data becomes available, using the framework we develop herein, can provide important insights into the effects of these changes and guide policy measures in future.

⁶¹ *Uniloc*, 632 F. 3d at 1292, 1335.

⁶² *Cornell*, 609 F. Supp. 2d at 279.

⁶³ *ResQNet.com*, 594 F.3d at 860.

⁶⁴ *Lucent*, 580 F.3d at 1301.

VI. Appendices

Appendix A

Dataset and Methodology:

Our dataset comprises a vast set of over 1000 U.S. District Court cases decided from 2004 to 2011, including summary judgments, bench trials, and jury verdicts. We exclude default judgments and other dismissals, as these are not representative of the complexity of most proceedings, and we also exclude cases primarily involving design patents, as the standards for design patent construction and infringement are considerably different than utility patents.⁶⁵

We start from a database of patent decisions maintained by PricewaterhouseCoopers (PwC). PwC uses this data in its annual reports on patent litigation, which are widely cited and used by academics, practitioners, and government policy-makers.⁶⁶ Working from the PwC dataset, we excluded dismissals and design patent cases, as well as a certain small proportion of cases where records were not accessible, yielding a dataset of 984 cases during this period. The figures below provide the breakdown of cases by year of decision, outcome (patent holder wins versus losses) and type of disposition (infringement versus invalidity versus non-infringement versus unenforceability).

Figure 16 below shows the number of cases by year of decision contained in the underlying PwC dataset, and Figure 17 provides a breakdown of patent holder wins versus accused infringer wins in each year.

⁶⁵ Specifically, we excluded cases where all or the majority of patents at issue in the case were design patents.

⁶⁶ PRICEWATERHOUSECOOPERS, 2015 PATENT LITIGATION STUDY, *available at* <https://perma.cc/4CN5-NYRE>.

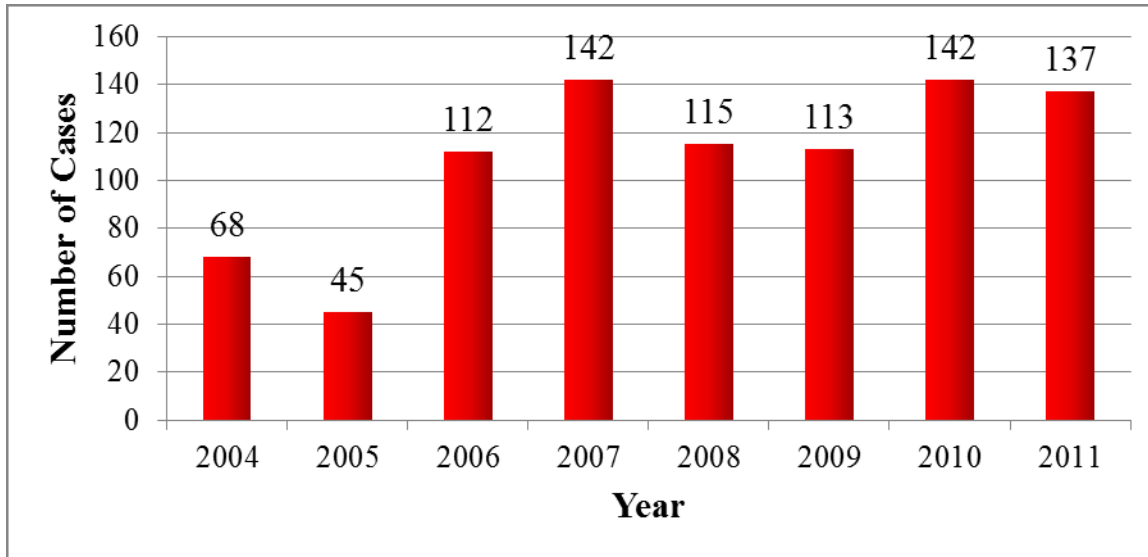


Figure 16: Number of Cases per Year

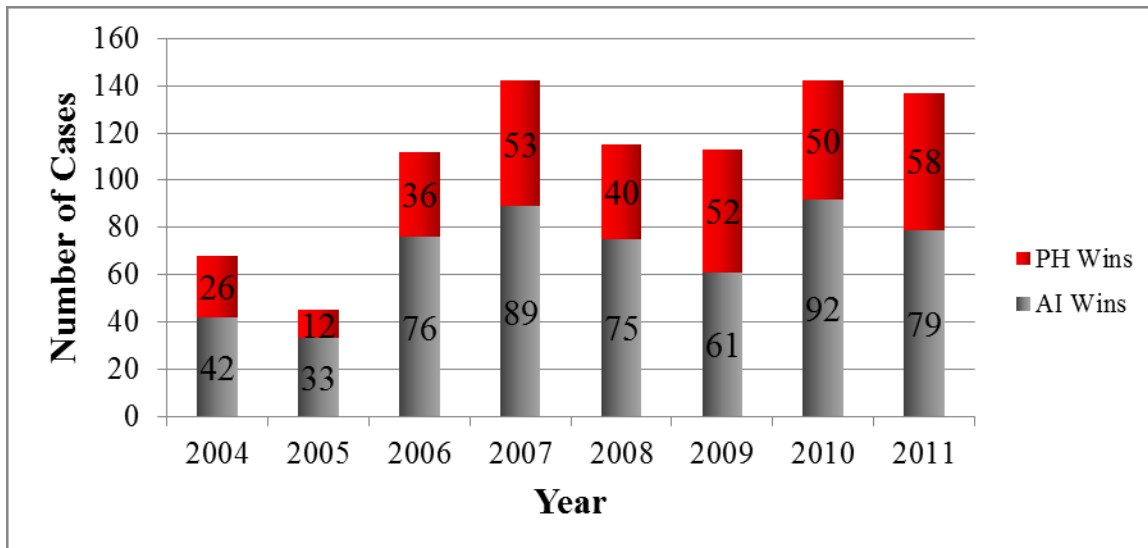


Figure 17: Number of PH Wins vs. AI Wins per Year

Finally, Figure 18 provides the breakdown of dispositions on each asserted patent in our dataset.

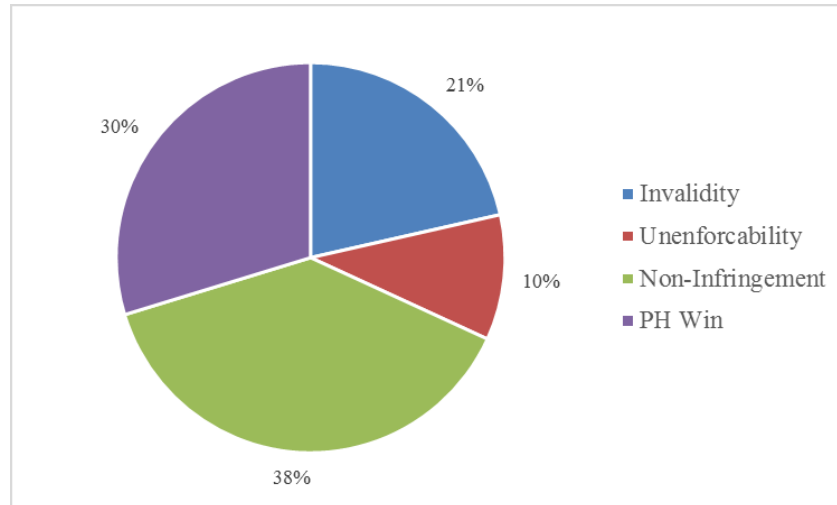


Figure 18: Breakdown of Cases by Type of Disposition

Appendix B-1

Case, Party and Patent Characteristics:

For each case, we coded over 100 variables describing characteristics of the cases and procedural posture, the individual litigants, and the patents at issue. Below is a summary of the principal variables we coded and details of our research procedure. A full list of variables follows in Appendix B-2.

- Case Variables: Our case variables include the particular U.S. District Court that heard the case as well as the Circuit in which such court was located. We recorded procedural details about each litigation, such as whether it was a declaratory judgment action, whether the case was decided on summary judgment or after a trial, and whether a bench or jury trial was held.

We read the initial complaint to identify the types of claims that were asserted. Where the case involved other claims in addition to infringement, we coded Boolean flags to denote the allegations, such as breach of contract (*e.g.*, in a patent licensing dispute), trade secret misappropriation, or antitrust or patent misuse claims. We read the opinions to determine whether these allegations were fully litigated or dropped along the way. We identified whether each case involved a claim of infringement based on filing of an Abbreviated New Drug Application (ANDA).

We used the dockets and opinions to identify whether a separate *Markman* hearing was conducted to construe the claims, whether a party moved for interlocutory appeal following the *Markman*, and whether such motion was granted. We recorded whether the Court granted a preliminary injunction or permanent injunction. We further identified venue transfers and stays where this was apparent from the dockets, as well as filings of appeal following the final judgment and other post-determination filings such as motions for vacatur or remittitur.

- Litigant Characteristics: We counted the number of plaintiffs and defendants in each case and recorded their names. We conducted research to determine whether they are publicly traded and where it was available, recorded their market capitalization and principal industry SIC codes. We also included

PwC's coding of whether the patent holder was a non-practicing entity and the specific type: an individual, university, or company (which includes Patent Assertion Entities). We further coded the number of law firms and attorneys of record representing each party.

- Patent Attributes: Finally, we used the Thomson Innovation patent databases to incorporate attributes of the patents at issue.⁶⁷ We recorded the application date and issue date (from which we computed the prosecution time and patent age at the date of the complaint), as well as the earliest priority date of each patent. We recorded the primary IPC code, whether the patent was a design or utility patent, and whether the patent had a PCT number representing an international filing. We coded the number of inventors, number of backward citations (broken down by patent and non-patent literature) and forward citations.⁶⁸ We also obtained the number of related patents and applications in the family tree of each asserted patent. We created Boolean flags indicating whether the patent had been reexamined, reissued or corrected (via a certificate of correction). Finally, we recorded the total number of claims and further auto-parsed the claim language to identify the total number of independent versus dependent claims.
- Pre-Processing Methodology: Certain of the raw data was converted into Boolean flags or grouped into categorical variables to avoid small bucket sizes or highly-skewed data. For example, patent IPC codes were categorized into 8 groupings based on the first letter industry marker, and SIC codes were

⁶⁷ Note that by "patents at issue" we are referring to the patents asserted in the original complaint. We also recorded the patents involved in the final dispositions, which in most cases were the same patents originally asserted. Even where some patents were dismissed or invalidated along the way, the process of doing so presumably may have contributed to the litigation duration and number of docket entries, and therefore we considered the patents asserted to be the most appropriate set for analyzing patent attributes.

⁶⁸ Given the age of the patents in these cases at the time of coding, we did not age-adjust forward citations (*e.g.*, using the NBER adjustment factors based on Hall, Jaffe & Trajtenberg's methodology). Rather, we used the current forward citation count as the estimate of lifetime citations, on the basis that most of these patents should have already received the vast majority of their citations. The average age at the time asserted is five years, and given case durations and the decision years in our dataset, nearly all patents are at least ten years old as of our coding.

grouped into 10 categories based on NAICS classification ranges according to the first two digits thereof. We also coded a Boolean flag to represent whether a single or multiple patent holder was named in the case and whether the patent holder(s) were represented by multiple law firms. We converted the total number of recorded assignments into a Boolean flag indicating whether or not the asserted patent had been assigned.

- In cases where multiple patents were at issue, we combined relevant attributes by computing averages and minimum/maximum values. For example, if a case involved three patents, patent A issued in 2001 and 1.2 years of age at the time of the complaint, patent B issued in 2002 and 2.5 years of age at the time of the complaint, and patent C issued in 2003 and 3.3 years of age at the time of the complaint, we used the average age of 2.3 years and minimum and maximum issue years of 2001 and 2003, respectively. If a patent in IPC A (Human Necessities) and a patent in IPC C (Chemistry) was asserted in the same case, we coded both flags as true. Similarly, if any of the asserted patents had been reexamined or corrected we coded the aggregate case flags as true.⁶⁹

For forward citations and backward citations, we calculated the averages of each of these quantities and recorded the maximum for all patents at issue in the case. We did not use the aggregate total of these fields to avoid double-counting—for example, for cases involving multiple patents by the same applicant, we expect some overlap in the citations made to previous patents as well as an increased likelihood in overlap in the forward citations received by each. Conversely, we recorded the average number of claims as well as the aggregate total number of claims across all patents asserted, as infringement and invalidity are claim-specific analyses and patent prosecution requirements impose limits on covering the same subject matter in multiple claims. Therefore, we

⁶⁹ In instances where multiple patents were at issue but one or more of them were missing certain fields in the Thomson Innovation databases, we computed the averages using the available data and reduced the averaging denominator to avoid reduction in the resulting quantity. For example, if the filing date of one of three patents was not available, we calculated the prosecution time (filing to issue) of the remaining two patents and used the average of these two times.

would expect each claim to be distinct, and each could potentially contribute to the overall complexity of the case.

- We also conducted testing to avoid cross-correlations and multi-collinearity in the data. Where two variables were strongly correlated we selected only one for the regression models, and where a set of variables exhibited multi-collinearity we dropped one or more.⁷⁰ We further constructed the final regression models via an iterative process, starting with a small number of unique variables and gradually adding additional independent variables and checking for significant changes in the resulting fit and degrees of freedom.⁷¹
- Finally, we log-transformed (natural logarithm) each of our complexity metrics to facilitate significance testing and regression analysis.

⁷⁰ For example, the minimum application year of the asserted patents was significantly negatively correlated with average patent age, and the resulting regression models contained significant oppositely-signed correlations for both variables. We excluded minimum application year from the to remove this effect and found that average patent age was not significant at the 5 percent level in the resulting model.

⁷¹ Multiple iterative ANOVA tests were conducted to check for significant changes between the different models.

Appendix B-2

Full List of Variables:

<u>Variable</u>	<u>Description</u>	<u>Source</u>
Case Info		
Case_ID	Unique case identifier (use this number at start of each associated filename).	Auto
P1_name	Name of captioned plaintiff.	Docket
D1_name	Name of captioned defendant.	Docket
Case_No	CV case number.	Docket
Dist_Ct	District court for this docket. (NOTE: Ask me if there was a venue transfer).	Docket
Compl_Dt	Date of initial complaint.	Docket
Dec_Year	Year of decision.	Docket
WL_cite	Westlaw citation (search party names, year and District if blank).	Docket
num_P	Number of plaintiffs.	Docket
num_D	Number of defendants.	Docket
P_names	Names of each plaintiff (separated by a semicolon).	Docket
D_names	Names of each defendant (separated by a semicolon).	Docket
DeclJ?	True if action against patent-holder for declaratory judgment of non-infringement.	Docket
Jury?	True if case decided by a jury; False for bench trial.	Docket
Trial?	True if case resulted in a trial.	Docket
Tr_Dt	Date of the trial (if any).	Docket
PH_Win?	True if patent-holder won (at least one patent held valid and infringed). NOTE: Will be False if plaintiff won in a DJ action seeking a declaration of non-infringement.	Opinion
Claim Info		
Breach_Pld?	True if pleadings also asserted a breach of contract claim.	Complaint/Answer
Misapp_Pld?	True if pleadings also asserted a misappropriation claim.	Complaint/Answer
Antitr_Pld?	True if pleadings also asserted an antitrust claim.	Complaint/Answer
Oth_Cl_Pld?	True if pleadings also asserted another claim (not listed above).	Complaint/Answer
Breach_Lit?	True if final opinion also adjudicated a breach of contract claim.	Opinion
Misapp_Lit?	True if final opinion also adjudicated a misappropriation claim.	Opinion
Antitr_Lit?	True if final opinion also adjudicated an antitrust claim.	Opinion
Oth_Cl_Lit?	True if final opinion also adjudicated another claim (not listed above).	Opinion
Outcome Info		
Opin?	True if District Court issued a written opinion.	Docket
SumJ?	True if case concluded by summary judgment.	Docket

Dism?	True if case concluded by dismissal.	Docket
ANDA?	True if infringement allegation is based on an ANDA filing.	Complaint/Opinion
Dsgn?	True if case included a design patent.	Complaint/Opinion
PermInj?	True if a permanent injunction was issued.	Docket
PreInj?	True if a preliminary injunction or TRO was issued.	Docket
inval?	True if any patent was held invalid.	Opinion
unenf?	True if any patent was held unenforceable.	Opinion
P_assrt	Patent numbers asserted.	Complaint
P_vald	Patent numbers held to be valid (separated by semicolons).	Opinion
P_inval	Patent numbers held to be invalid (separated by semicolons).	Opinion
P_enf	Patent numbers held to be enforceable (separated by semicolons).	Opinion
P_unenf	Patent numbers held to be unenforceable (separated by semicolons).	Opinion
P_infr	Patent numbers held infringed (separated by semicolons).	Opinion
P_noinfr	Patent numbers held non-infringed (separated by semicolons).	Opinion
Def_Jdgmt	Defendants (or Plaintiffs in a DJ action) who were involved in the final opinion/order. (Names separated by semicolons.)	Opinion
<i>Docket Info</i>		
Num_DocEv_D Ct	Number of docket events from the complaint to the final decision (including remittitur/vacatur but not including subsequent appeal).	Docket
Num_DocEv_To t	Total number of docket events.	Docket
Num_P_firms	Number of law firms representing plaintiffs.	Docket
Num_D_firms	Number of law firms representing defendants.	Docket
P_firm_names	Names of law firms representing plaintiffs (separated by semicolons).	Docket
D_firm_names	Names of law firms representing defendants (separated by semicolons).	Docket
Num_P_atty	Number of named attorneys representing plaintiffs.	Docket
Num_D_atty	Number of named attorneys representing defendants.	Docket
Num_P_filings	Number of *total filings* filed by plaintiff(s) (excluding appeal).	Docket
Num_P_mo	Number of motions filed by plaintiff(s) (excluding appeal).	Docket
Num_D_filings	Number of *total filings* filed by defendant(s) (excluding appeal).	Docket
Num_D_mo	Number of motions filed by defendant(s) (excluding appeal).	Docket
Num_Ct_filings	Number of *total filings* filed by court (excluding appeal).	Docket
Num_Ct_ord	Number of memoranda/opinions/orders by the court	Docket

	(excluding appeal).	
Markman?	True if there was a Markman hearing for claim construction.	Docket
Interl_App_M?	True if a party FILED for interlocutory appeal after the Markman.	Docket
Interl_App_G?	True if the court GRANTED motion for interlocutory appeal after the Markman. (NOTE: Ask me if there was an interlocutory appeal at another point in the case.)	Docket
PreM_P_filings	Number of *total filings* filed by plaintiff(s) pre-Markman (if applicable).	Docket
PreM_P_mo	Number of motions filed by plaintiff(s) pre-Markman (if applicable).	Docket
PreM_D_filings	Number of *total filings* filed by defendant(s) pre-Markman (if applicable).	Docket
PreM_D_mo	Number of motions filed by defendant(s) pre-Markman (if applicable).	Docket
PreM_Ct_filings	Number of *total filings* filed by court pre-Markman (if applicable).	Docket
PreM_Ct_ord	Number of memoranda/opinions/orders by the court pre-Markman (if applicable).	Docket
Ven_Tr?	True if there was a venue transfer (include details in Notes).	Docket
Stay?	True if the litigation was stayed at any point (include details in Notes).	Docket
Num_Amici	Number of amici briefs filed with the court (if any).	Docket
Rem_Vac_M?	True if party FILED for remittitur or vacatur post-decision.	Docket
Rem_Vac_G?	True if the court GRANTED motion for remittitur or vacatur post-decision.	Docket
Appeal_M?	True if a party FILED for appeal.	Docket
Party Info		
P_Public?	True if Plaintiff is a public company.	MergentOnline
P_cap	Plaintiff's market capitalization / private valuation.	MergentOnline
P_IndSIC	4-digit SIC code of Plaintiff.	MergentOnline
P_NPE?	True if Plaintiff is a non-practicing entity.	PWC
D_Public?	True if Defendant is a public company.	Hoover/Mergent
D_cap	Defendant's market capitalization / private valuation.	Hoover/Mergent
Patent Info		
Application		
Date	Application date of the patent.	Thomson Innovation
Priority Date - Earliest	Earliest priority date of the patent.	Thomson Innovation
PriorPar?	Does the patent claim priority from an earlier application? (T/F)	Thomson Innovation
Issue Date	Issue date of the patent.	Thomson Innovation
ProsecTime	Duration between application and issue.	Calculated
AgeAtCompl	Duration between issue and complaint.	Calculated
PCT?	Does the patent have a PCT number? (T/F)	Thomson Innovation

IPC - Current	IPC Codes of the patent.	Thomson Innovation
US Class	US Classification Codes of the patent.	Thomson Innovation
NmOrigAssg?	Does the patent name an original assignee? (T/F)	Thomson Innovation
Assigned?	Has the patent been assigned (based on USPTO records)? (T/F)	Thomson Innovation
Inventor Count	Number of named inventors of the patent.	Thomson Innovation
BC_Pat	Number of backward citations to patent references.	Thomson Innovation
BC_Lit	Number of backward citations to non-patent references.	Thomson Innovation
FC	Number of forward citations.	Thomson Innovation
FamilySize	Size of the patent family of which this patent is a member.	Thomson Innovation
Reiss?	Was the patent reissued?	Thomson Innovation
Reex?	Was the patent reexamined?	Thomson Innovation
Corr?	Was the patent corrected?	Thomson Innovation
NumCl	Total number of claims.	Thomson Innovation
NumIndep	Number of independent claims.	Thomson Innovation

Appendix C

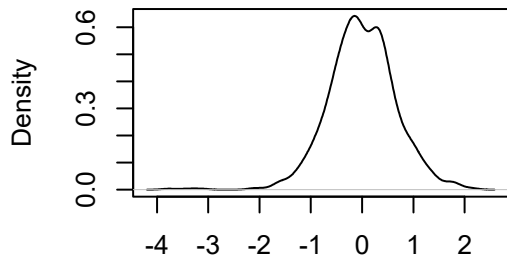
Regression Results—Docket Entries to Disposition:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
<i>Case, Disposition and Procedure Variables</i>					
• Year of Disposition	0.04	0.01	3.09	0.00	**
• 1 st Circuit	0.16	0.23	0.68	0.50	
• 3 rd Circuit	0.55	0.21	2.62	0.01	**
• 5 th Circuit	0.22	0.21	1.03	0.30	
• 7 th Circuit	0.50	0.22	2.33	0.02	*
• 9 th Circuit	0.50	0.21	2.38	0.02	*
• ANDA?	-0.11	0.11	-1.02	0.31	
• Invalid Patent?	-0.01	0.06	-0.16	0.87	
• Unenforceable Patent?	-0.08	0.09	-0.89	0.37	
• Non-Infringed Patent?	-0.13	0.06	-2.16	0.03	*
• Venue Transfer?	0.01	0.12	0.11	0.91	
• Stay?	0.20	0.05	3.60	0.00	***
• Markman?	0.34	0.05	6.92	0.00	***
• Interloc. Appeal?	0.43	0.10	4.35	0.00	***
• Jury Trial?	0.61	0.07	8.19	0.00	***
• Bench Trial?	0.16	0.07	2.41	0.02	*
<i>Litigant Variables</i>					
• # Accused Infringers	0.04	0.01	4.85	0.00	***
• # AI Firms	0.06	0.01	5.12	0.00	***
• Large-Entity AI?	0.09	0.06	1.63	0.10	
• Multiple Patent Holders?	0.13	0.05	2.52	0.01	*
• Multiple PH Firms?	0.12	0.05	2.28	0.02	*
• Large-Entity PH?	0.18	0.06	2.90	0.00	**
• NPE (Individual)	-0.14	0.09	-1.67	0.09	.
• NPE (Company)	0.06	0.08	0.81	0.42	
• NPE (University)	-0.35	0.30	-1.16	0.24	
<i>Patent Variables</i>					
• Avg. FC	0.00	0.00	0.48	0.63	
• Avg. Age	0.00	0.00	-0.10	0.92	
• Max. Family Size	0.01	0.00	3.38	0.00	***
• Multiple Patents?	0.08	0.05	1.50	0.13	
• IPC A?	-0.10	0.07	-1.45	0.15	
• IPC B?	-0.01	0.08	-0.18	0.86	
• IPC G or H?	-0.06	0.07	-0.88	0.38	
• High-Tech?	0.14	0.07	2.03	0.04	*

Full regression results on file with the author.

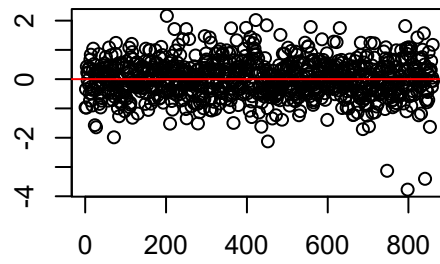
R²: 0.460
Adj.R²: 0.427
Std.Err: 0.697
F_(55,911): 14.09
p-val: 2.2e-16
N: 967

Distribution of Residuals

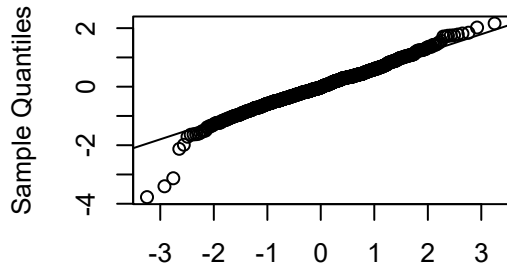


N = 861 Bandwidth = 0.1402

Residuals Plot



Normal Q-Q Plot



Residuals Plots of Docket Entries to Disposition Regression

Appendix D

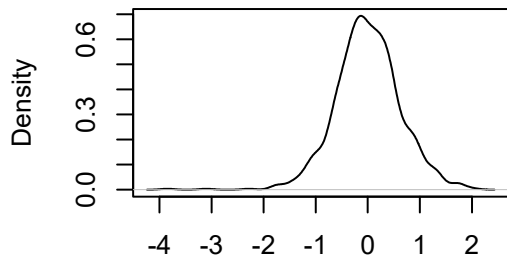
Regression Results—Total Docket Entries:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
<i>Case, Disposition and Procedure Variables</i>					
• Year of Disposition	0.03	0.01	2.19	0.03	*
• 1 st Circuit	0.29	0.22	1.29	0.20	
• 3 rd Circuit	0.61	0.21	2.96	0.00	**
• 5 th Circuit	0.25	0.21	1.21	0.23	
• 7 th Circuit	0.56	0.21	2.68	0.01	**
• 9 th Circuit	0.55	0.20	2.69	0.01	**
• ANDA?	-0.11	0.11	-1.02	0.31	
• Invalid Patent?	0.06	0.06	0.93	0.36	
• Unenforceable Patent?	-0.08	0.09	-0.92	0.36	
• Non-Infringed Patent?	-0.07	0.06	-1.26	0.21	
• Venue Transfer?	-0.01	0.12	-0.09	0.93	
• Stay?	0.22	0.05	4.17	0.00	***
• Markman?	0.34	0.05	7.05	0.00	***
• Interloc. Appeal?	0.37	0.10	3.90	0.00	***
• Jury Trial?	0.62	0.07	8.61	< 2e-16	***
• Bench Trial?	0.12	0.07	1.79	0.07	.
<i>Litigant Variables</i>					
• # Accused Infringers	0.03	0.01	4.98	0.00	***
• # AI Firms	0.06	0.01	5.22	0.00	***
• Large-Entity AI?	0.11	0.05	1.95	0.05	.
• Multiple Patent Holders?	0.13	0.05	2.59	0.01	**
• Multiple PH Firms?	0.16	0.05	3.17	0.00	**
• Large-Entity PH?	0.12	0.06	2.03	0.04	*
• NPE (Individual)	-0.17	0.08	-2.11	0.04	*
• NPE (Company)	0.04	0.08	0.58	0.56	
• NPE (University)	-0.05	0.29	-0.17	0.87	
<i>Patent Variables</i>					
• Avg. FC	0.00	0.00	0.25	0.80	
• Avg. Age	0.00	0.00	0.10	0.92	
• Max. Family Size	0.01	0.00	3.78	0.00	***
• Multiple Patents?	0.05	0.05	1.03	0.31	
• IPC A?	-0.09	0.07	-1.40	0.16	
• IPC B?	-0.02	0.08	-0.21	0.83	
• IPC G or H?	-0.07	0.07	-1.03	0.30	
• High-Tech?	0.14	0.07	2.16	0.03	*

Full regression results on file with the author.

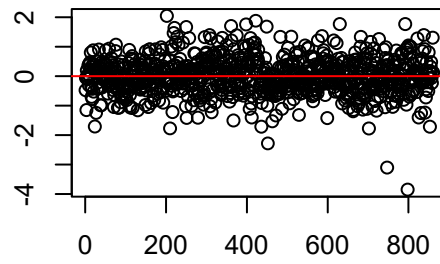
R²: 0.477
 Adj. R²: 0.445
 Std.Err.: 0.674
 F_(55,911): 15.09
 p-val: 2.2e-16
 N: 967

Distribution of Residuals

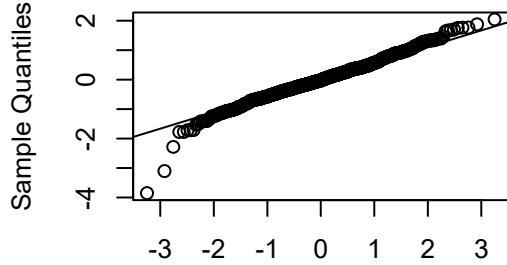


N = 861 Bandwidth = 0.1303

Residuals Plot



Normal Q-Q Plot



Residuals Plots of Total Docket Entries Regression

Appendix E

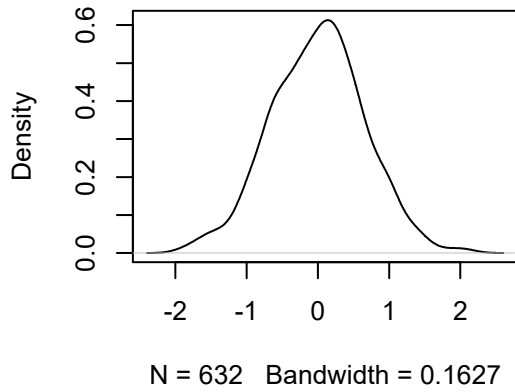
Regression Results—Substantive Docket Entries to Disposition:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
<i>Case, Disposition and Procedure Variables</i>					
• Year of Disposition	0.01	0.02	0.31	0.76	
• 1 st Circuit	-0.07	0.27	-0.25	0.80	
• 3 rd Circuit	0.09	0.24	0.38	0.71	
• 5 th Circuit	0.21	0.25	0.86	0.39	
• 7 th Circuit	0.14	0.25	0.55	0.58	
• 9 th Circuit	0.09	0.24	0.39	0.70	
• ANDA?	-0.13	0.14	-0.92	0.36	
• Invalid Patent?	0.09	0.08	1.13	0.26	
• Unenforceable Patent?	0.15	0.14	1.01	0.31	
• Non-Infringed Patent?	-0.03	0.08	-0.38	0.70	
• Venue Transfer?	-0.03	0.14	-0.25	0.80	
• Stay?	0.23	0.07	3.36	0.00	***
• Markman?	0.40	0.06	6.42	0.00	***
• Interloc. Appeal?	0.42	0.12	3.48	0.00	***
• Jury Trial?	0.39	0.10	4.10	0.00	***
• Bench Trial?	0.20	0.10	2.05	0.04	*
<i>Litigant Variables</i>					
• # Accused Infringers	0.04	0.01	3.38	0.00	***
• # AI Firms	0.05	0.01	3.87	0.00	***
• Large-Entity AI?	0.10	0.07	1.46	0.14	
• Multiple Patent Holders?	0.08	0.06	1.32	0.19	
• Multiple PH Firms?	0.19	0.07	2.85	0.00	**
• Large-Entity PH?	0.21	0.08	2.77	0.01	**
• NPE (Individual)	0.00	0.10	-0.02	0.98	
• NPE (Company)	0.15	0.10	1.53	0.13	
• NPE (University)	0.28	0.36	0.76	0.45	
<i>Patent Variables</i>					
• Avg. FC	0.00	0.00	0.03	0.97	
• Avg. Age	0.00	0.00	-0.44	0.66	
• Max. Family Size	0.00	0.00	1.04	0.30	
• Multiple Patents?	0.04	0.07	0.62	0.54	
• IPC A?	-0.04	0.08	-0.45	0.65	
• IPC B?	0.08	0.10	0.84	0.40	
• IPC G or H?	-0.07	0.09	-0.87	0.39	
• High-Tech?	0.09	0.08	1.14	0.26	

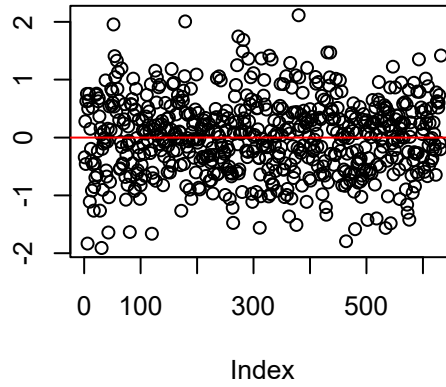
Full regression results on file with the author.

R²: 0.455
 Adj. R²: 0.403
 Std.Err.: 0.686
 F_(55,576): 8.74
 p-val: 2.2e-16
 N: 632

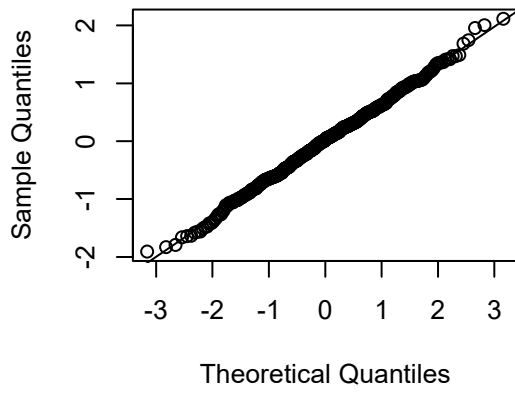
Distribution of Residuals



Residuals Plot



Normal Q-Q Plot



Residuals Plots of Substantive Docket Entries to Disposition Regression

Appendix F

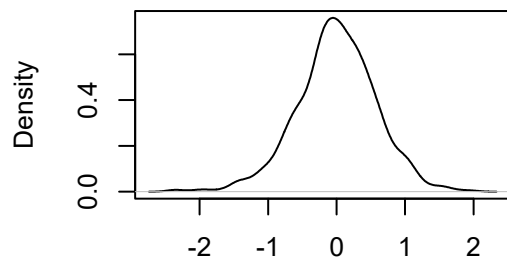
Regression Results—Duration to Disposition:

<i>Regressor</i>	<i>Coeff.</i>	<i>Std.Err.</i>	<i>t-val</i>	<i>Pr(> t)</i>	<i>Signif.</i>
<i>Case, Disposition and Procedure Variables</i>					
• Year of Disposition	0.01	0.01	0.60	0.55	
• 1 st Circuit	0.05	0.20	0.23	0.82	
• 3 rd Circuit	0.05	0.19	0.24	0.81	
• 5 th Circuit	-0.17	0.19	-0.92	0.36	
• 7 th Circuit	-0.10	0.19	-0.55	0.58	
• 9 th Circuit	-0.15	0.19	-0.81	0.42	
• ANDA?	-0.06	0.10	-0.62	0.53	
• Invalid Patent?	0.06	0.06	1.05	0.29	
• Unenforceable Patent?	-0.09	0.08	-1.08	0.28	
• Non-Infringed Patent?	0.09	0.05	1.79	0.07	.
• Venue Transfer?	0.19	0.11	1.71	0.09	.
• Stay?	0.21	0.05	4.34	0.00	***
• Markman?	0.14	0.04	3.16	0.00	**
• Interloc. Appeal?	0.26	0.09	3.01	0.00	**
• Jury Trial?	-0.01	0.07	-0.08	0.93	
• Bench Trial?	0.03	0.06	0.58	0.57	
<i>Litigant Variables</i>					
• # Accused Infringers	0.00	0.01	-0.40	0.69	
• # AI Firms	0.02	0.01	2.21	0.03	*
• Large-Entity AI?	-0.10	0.05	-2.06	0.04	*
• Multiple Patent Holders?	0.06	0.05	1.36	0.17	
• Multiple PH Firms?	0.03	0.05	0.74	0.46	
• Large-Entity PH?	0.03	0.06	0.60	0.55	
• NPE (Individual)	0.17	0.07	2.24	0.03	*
• NPE (Company)	0.12	0.07	1.81	0.07	.
• NPE (University)	-0.44	0.26	-1.68	0.09	.
<i>Patent Variables</i>					
• Avg. FC	0.00	0.00	1.19	0.23	
• Avg. Age	0.00	0.00	-1.71	0.09	.
• Max. Family Size	0.00	0.00	1.69	0.09	.
• Multiple Patents?	-0.01	0.05	-0.23	0.82	
• IPC A?	0.04	0.06	0.69	0.49	
• IPC B?	0.13	0.07	1.95	0.05	.
• IPC G or H?	-0.01	0.06	-0.17	0.86	
• High-Tech?	-0.04	0.06	-0.58	0.56	

Full regression results on file with the author.

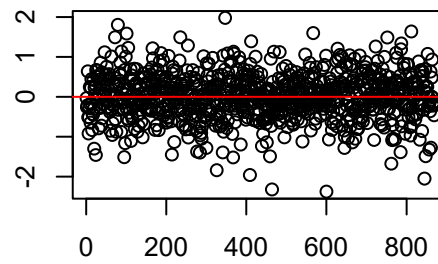
R²: 0.196
 Adj. R²: 0.149
 Std.Err.: 0.616
 F_(55,921): 4.09
 p-val: 2.2e-16
 N: 977

Distribution of Residuals

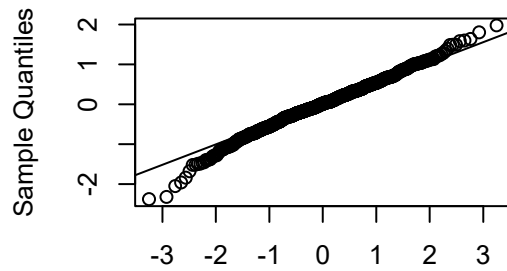


N = 869 Bandwidth = 0.12

Residuals Plot



Normal Q-Q Plot



Theoretical Quantiles

Residuals Plots of Duration to Disposition Regression