PREDICTING THE "UNPREDICTABLE": AN EMPIRICAL ANALYSIS OF U.S. PATENT INFRINGEMENT AWARDS

Michael J. Mazzeo
Kellogg School of Management, Northwestern University

Jonathan Hillel
Skadden, Arps, Slate, Meagher & Flom LLP

Samantha Zyontz
Harvard Business School

ABSTRACT:

Patent infringement awards are commonly feared to be unpredictable, and such unpredictability is decried as a significant problem in the patent system. We investigate the assumption that patent damages are unpredictable by conducting a large-scale econometric analysis of award values. We find a high degree of correlation between award value and *ex ante*-identifiable factors. We begin by analyzing the outcomes of 340 cases decided in US federal courts between 1995 and 2008 in which infringement was found and damages were awarded. Our data include the amount awarded, along with information about the litigants, case specifics and economic value of the patents-at-issue. Using these data, we construct an econometric model that explains over 75% of the variation in awards. We further conduct in-depth analysis of the key factors affecting award value, via targeted regressions involving fewer variables. Our findings refute claims that infringement awards are systematically unpredictable and, moreover, highlight the critical elements that can be expected to result in larger or smaller awards.

PREDICTING THE "UNPREDICTABLE": AN EMPIRICAL ANALYSIS OF U.S. PATENT INFRINGEMENT AWARDS

Michael J. Mazzeo[†*]
Jonathan Hillel[**]
Samantha Zyontz[***]

INTRODUCTION

Patent infringement awards are commonly feared to be unpredictable. Patents are often characterized as volatile assets with the potential to give rise to blockbuster awards and "bet-the-company" liabilities. The most recent Federal Trade Commission report on the patent system highlights a "lottery ticket mentality" towards patent litigation outcomes.[1] Furthermore, the huge amounts spent on prosecuting and defending patent cases, assessed by the AIPLA to average $6.25 million in cases with over $25 million at issue, reflect both the expected stakes of patent litigation and the unpredictability of the resulting outcomes.[2]

The assumption of unpredictability has also pervaded the policy debate and given rise to several reform measures. Before passage of the America Invents Act, the leading damages proposal sought to bolster the judge's role as the "gatekeeper" of evidence,[3] with the explicit aim

---

† The authors are grateful foremost to F. Scott Kieff and Geoffrey J. Lysaught for their insights and comments during early development of previous outlines. The authors are also grateful to Christopher Barry, Vincent Torres, and Erik Skramstad at PricewaterhouseCoopers LLP for supplying the proprietary database on which this study was built. We also thank Josh Wright, David Schwartz, Max Schanzenbach, Henry Butler, the Searle Civil Justice Institute Board of Overseers, the Searle Civil Justice Institute Faculty Advisory Committee and participants at the Samsung-Stanford Conference on Patent Remedies, the Patent Conference at the University of Kansas and at the workshops at George Mason University Law School, University of Rochester, and Indiana University and Mark Schankerman, Jason Du Mont and the participants at the Max Planck Institute's Third Workshop for Junior Researchers on the Law & Economics of Intellectual Property and Competition Law for their helpful comments on previous outlines and drafts. Elise Nelson and Matthew Sibery must also be thanked for their tireless research assistance.
The authors are grateful for the generous support of the Searle Center at Northwestern University School of Law and Microsoft Corporation.
The views expressed are solely those of the authors and do not reflect the views of others, including PricewaterhouseCoopers LLP, Northwestern University, Microsoft Corporation, Skadden, Arps, Slate, Meagher & Flom LLP (or its attorneys or clients), or any of their affiliates.

* Associate Professor of Management & Strategy, Kellogg School of Management, Northwestern University and Faculty Associate, Institute for Policy Research, Northwestern University.
** Associate, Skadden, Arps, Slate, Meagher & Flom LLP.
*** Research Manager, Institute for Strategy and Competitiveness, Harvard Business School.

1 Federal Trade Commission, *The Evolving IP Marketplace, Aligning Patent Notice and Remedies with Competition* (March 2011), *available at* www.ftc.gov/os/2011/03/110307patentreport.pdf [hereinafter "2011 FTC Report"].

2 Am. Intell. Prop. Law Ass'n., REPORT OF THE ECONOMIC SURVEY I-131 (2009).

3 The "gatekeeper" proposal would have augmented the judge's role as evidentiary gatekeeper by requiring the judge to exclude all methodologies and factors used in calculating infringement damages that are not supported by "sufficient" evidence. *See* S. 515 § 4 (proposed amendment to 35 U.S.C. § 284(b)(1))

of preventing jury errors and runaway verdicts.[4] Furthermore, case law developments preceding and following passage of the AIA, directed at restricting evidentiary rules and limiting fact-finder discretion, are arguably intended to improve clarity and predictability of damage awards.[5]

This study challenges the assumption that patent damages are unpredictable. We study the predictability of patent infringement awards over a 14-year period via regression analysis. We find that *ex ante*-observable factors of the litigants, case specifics and economic value of the patents-at-issue explain over 75% of the variation of the resulting infringement awards. We further study the significant factors influencing award value and compare them to factors known to influence rates of patent litigation.

In our study, we systematically catalogue the size of damage awards and explore factors that contribute to the observed dollar amount of awards. We analyze 340 patent infringement damage awards granted by a judge or jury in United States district courts from 1995 to 2008, using the economic value of patents as a benchmark.[6] These data were derived from a proprietary dataset owned by PricewaterhouseCoopers ("PwC") and licensed to us for use in this study. The PwC dataset, which informed the proposed patent reform legislation,[7] contains over 1,300 final patent decisions in US district courts from 1995 to 2008. We supplement the PwC dataset by reviewing the original case records for data regarding the damages theories used, asserted patents, procedural disposition, as well as venue and party characteristics. We then code these data for over 120 variables describing various aspects of the cases and awards. We analyze these data using regression analysis, seeking in the first stage to achieve maximum R-squared fit to the data, and other standard statistical methods. The result is a comprehensive empirical evaluation of the nature and characteristics of patent infringement damage awards in US district courts during this 14 year period.[8]

Our key findings are as follows:

- The distribution of award levels is skewed, with a small number of very high dollar valued awards relative to the rest of the distribution. Specifically, the largest eight awards comprised over 47% of the aggregate awards amount over the time period studied.
- Infringement damages are highly predictable using the factors we included as explanatory variables. Our econometric model accounts for over 75% of variation across the dataset.

---

4 Senate Report on the patent reform Act of 2009, S. Rep. 111-18, at 8 (May 12, 2009) ("damage awards . . . are too often excessive and untethered from the harm that compensatory damages are intended to measure").

5 *See* Section I(C), *infra*.

6 We refer to the economic literature on patent valuation to build a statistical model based on factors that have been shown to affect the economic value of patents. *See* note 184, *infra*, and accompanying text.

7 *See*, *e.g.*, 2009 Senate Report, *supra* note 1, at 9 n.40 (*citing* 2007 PwC Study).

8 Our analysis may miss some patent infringement damage awards from cases where relevant information was not reported (though we believe the impact on our conclusions to be minimal). Further, as the dataset only contains awards in US district courts before appeal, we cannot make definitive statements about the effect of the higher courts' decisions on final patent damage awards. Caveats regarding our findings are discussed further in our concluding section.

- Our analysis of significant factors influencing patent awards finds that the following tend to be associated with *higher* award values:
    - More patents per case;
    - More mature patents;
    - Patents with more claims and patents with more forward citations;
    - Patent-holders who manufacture and/or market the patented technology, as opposed to non-practicing entities;
    - Cases decided by juries; and
    - More complex cases (as measured by longer times to trial).

Section I analyzes the law of patent infringement damages, perceived problems and various proposals to address them, prior empirical studies of patent infringement damages, and recent relevant case law from the Federal Circuit and certain district courts. Section II outlines the research methodology employed in this article and presents descriptive statistics about the dataset. Section III provides the results of the empirical analysis. Finally, Section IV concludes by discussing policy implications and questions for future study.

I. BACKGROUND

This Section provides relevant background for our empirical analysis. Part A outlines the statutory and case law of patent infringement damages. Part B surveys previous empirical studies of patent infringement damages. Part C discusses recent federal case law regarding damages, and finally Part D explores certain implications of these decisions.

A. Summary of the Law of Patent Infringement Damages

A patent confers the right to exclude others from making, using, selling, offering for sale or importing the invention defined by its claims.[9] Section 284 of the Patent Act of 1952 provides damages for infringement of patent rights. Pursuant to Section 284, a successful claimant is entitled to receive "damages adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer."[10] The statute affords no further explanation of the composition or calculation of compensatory damages, which has given rise to extensive litigation and a library of legal scholarship.

The two primary theories for awarding patent infringement damages are lost profits and reasonable royalties.[11] As its name suggests, lost profits awards the patentee[12] the profits that it lost as a result of the infringement.[13] To recover these damages, the claimant bears the burden of proving it is entitled to, and the amount of, lost profits.[14] By contrast, a claimant is entitled to a reasonable royalty upon proof of infringement, but nonetheless bears the burden of proving its claimed amount of reasonable royalty damages by a preponderance of evidence.[15] Properly construed, a reasonable royalty is merely "the floor below which damages shall not fall."[16] A claimant may elect whether to proceed on a theory of lost profits or claim reasonable royalty damages without seeking lost profits.[17]

B. Previous Empirical Studies

---

9 35 U.S.C. § 271.

10 § 284.

11 Panduit Corp. v. Stahlin Bros. Fibre Works, Inc., 575 F.2d 1152 (6th Cir. 1978) (lost profits); Georgia-Pacific Corp. v. US Plywood Corp. 318 F. Supp. 1116, 1120 (S.D.N.Y. 1970) (reasonable royalties); *see also* State Indus., Inc. v. Mor-Flo Indus., Inc., 883 F.2d 1573 (Fed. Cir. 1989) (alternate methodologies); *see generally* RICHARD CAULEY, WINNING THE PATENT DAMAGES CASE: A LITIGATOR'S GUIDE TO ECONOMIC MODELS AND OTHER DAMAGES STRATEGIES (Oxford University Press 2008).

12 Note that "patentee," as used herein, refers to any party with standing to claim damages for patent infringement. This may include the original patent owner, assignees, or certain exclusive licensees.

13 JANICE M. MULLER, PATENT LAW 498 (3rd ed. 2009).

14 John M. Skenyon, Christopher S. Marchese & John Land, Patent Damages Law and Practice § 1:3 (Aug. 2008).

15 *Id.* § 1:3.

16 Bandag, Inc. v. Gerrard Tire Co., 704 F.2d 1578, 1583 (Fed. Cir. 1983).

17 *Id.*

Certain previous studies have undertaken large-scale analysis of patent damage awards.[18] Several studies by Lanjouw and Schankerman from 1999-2004 study the predictability and determinants of patent infringement suits.[19] A 2007 PricewaterhouseCoopers study (the "2007 PwC Study") finds a fivefold disparity between median jury verdicts and median bench awards.[20] A 2008 update to the study (the "2008 PwC Study") provides supplementary data and analysis.[21] A 2009 update to the study (the "2009 PwC Study") provides supplementary data and new analysis of the impact of nonpracticing entities ("NPEs") engaging in patent litigation.[22] A 2007 study by Lemley & Shapiro addresses reasonable royalty awards and apportionment in multi-component products.[23] One recent study by Allison, Lemley & Walker address patent litigation in different industry sectors, and find that litigation rates and litigant characteristics vary significantly by industry.[24] Finally, Opderbeck conducts empirical analysis that questions the assumption that patent infringement awards are systematically excessive.[25]

1. Lanjouw and Schankerman Studies

Lanjouw and Schankerman provide path-breaking analysis of the predictability of patent litigation and the litigation value of patents across several studies. In their 2001 study, they study the factors influencing litigation rates and win/loss/settlement outcomes. They identify certain patent characteristics lending to an increased likelihood of suit, most notably a higher number of claims and more forward citations per claim (so-called "valuable patents").[26] Additionally, certain litigant characteristics are found to influence litigation rates, such as the entity type and size of the patent owner.[27] They find that public companies are significantly less likely to file suits on patents they own than smaller, non-public companies and individuals.

2. 2007 PwC Study

The 2007 PwC Study aggregated bench awards and jury verdicts in the years 1980 to 2005. It contained two findings relevant to the present focus. First, it found that median jury

---

18 Many of these studies were cited in the 2011 FTC Report as the "available statistics on patent litigation outcomes and damage awards." 2011 FTC Report, *supra* note 88, at 162. As described below, significant data gaps exist in this precedent.

19 *See* Lanjouw, J. O. and Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, Rand J. Econ. Vol. 32, no. 1, pp. 129-51 (2001); Lanjouw, J. O. and Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, J. L. and Econ. Vol. XLVII, no. 1. pp. 45-74 (2004); Lanjouw, J. O. and Mark Schankerman, *Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators*, Econ. J.. Vol. 114, pp. 441-65 (2004).

20 PricewaterhouseCoopers, 2007 Patent and Trademark Damages Study [hereinafter "2007 PwC Study"].

21 PricewaterhouseCoopers, A Closer Look—Patent Litigation Study: Damage Awards, Success Rates and Time-To-Trial (2008) [hereinafter "2008 PwC Study"].

22 PricewaterhouseCoopers, A Closer Look—Patent Litigation Trends and the Increasing Impact of Nonpracticing Entities (2009) [hereinafter "2009 PwC Study"].

23 Lemley & Shapiro, *supra* note 50.

24 John R. Allison, Mark A. Lemley & J.H. Walker, *Trolls on Top?*, 158 U.Penn.L.Rev. 1 (studying litigation rates of patents in specific industries). However, this study does not address the outcomes of the litigation, but notes "that is the subject of a companion piece by the authors, tentatively entitled Patent Quality and Risk Aversion Among Repeat Patent Litigants." *Id.* at 5 n. 14.

25 David W. Opderbeck, *Patent Damages Reform and the Shape of Patent Law*, 89 B.U.L.Rev. 127 (2009).

26 [_]

27 [_]

awards were on average five times larger than median awards in bench trials, during the years studied.[28] Second, the study provided trend data on royalty rates, finding that average awarded royalty rates have declined in recent years.[29]

The analytic methodology utilized in the 2007 PwC Study is as follows:[30]

- "PricewaterhouseCoopers identified legal records in two Westlaw databases, Federal Intellectual Property—Cases (FIP-CS) and Combined Jury Verdicts and Settlements (JV-ALL), from 1980 through June 2006."
- "The study included . . . [1,367] unique US Federal District Court [patent] cases . . . and 29 cases that included both patent and trademark issues . . . and [273] unique CAFC [patent] cases . . . 7 cases that included both patent and trademark issues."
- "Jury verdict information varied by jurisdiction and was particularly limited during the early and mid-1980s."

3. 2008 PwC Study

The 2008 PwC Study contains an updated dataset, containing data on damage awards from 1980-2007.[31] Furthermore, additional analysis is provided. Notably, the study lists nine "landmark" awards from 2005-2007 that exceeded $100M. The study also lists six of the largest awards since 1985.[32]

The 2008 PwC Study considers the incidence of bench versus jury decisions, finding "a marked increase in jury trials since the 1980s, with the shift becoming more evident since 1995."[33] The study attributes this increase to a "stark contrast" in plaintiff success rates between bench and jury trials, and median jury awards that are "significantly larger" than median bench awards. The study finds that "[j]ury success rates have consistently outperformed their bench counterparts for every year since 1995."[34] Additionally, "[r]ecent awards by juries have been running several multiples of the amounts awarded by judges."[35]

The 2008 PwC Study further performs some initial industry-specific analysis. For instances, the study calculates the median damages award in ten industry sectors.[36] The authors do not explicitly describe their methodology for identifying the industry sectors. The study also ranks judicial districts according to median damage awards from 1995 to 2007. It finds that "[c]ertain federal district courts . . . continue to be more favorable to patent holders."[37] Finally, the 2008 PwC Study considers appeal rates and appellate outcomes (affirmance, reversal, or modification).

---

28 2007 PwC Study, *supra* note 95, at 14.
29 *Id.* at 22-25.
30 *Id.* at 29.
31 2008 PwC Study, *supra* note 96, at 1.
32 *Id.* at 3-4.
33 *Id.* at 4.
34 *Id.* at 5.
35 *Id.* at 6.
36 *Id.* at 3.
37 *Id.* at 14.

4. 2009 PwC Study

The 2009 PwC Study contains an updated dataset, including information on damage awards from 1980 to 2008.[38] Furthermore, new analysis is conducted on NPEs involved in patent litigation. The study defines an NPE as "an entity that does not have the capabilities to design, manufacture, or distribute products that have features protected by the patent."[39]

Among its key findings, the 2009 PwC Study determined that the median patent infringement damages award for NPE patent-holders was more than three times that of practicing entities during the period from 2002 to 2009.[40] Whereas the median during this period for practicing entities was $3.4 million, it was $12 million for NPEs (in inflation-adjusted numbers); by contrast, from 1995 to 2001, the medians were roughly equal for NPEs and practicing entities alike.[41] Also, like the 2008 PwC Study, this iteration also lists the "landmark" awards from 2005-2007 that exceeded $100M, and further indicates the entity status of the patentee.[42]

In addition, the 2009 PwC Study reports the incidence of bench versus jury decisions and median bench versus jury damage awards categorized by type of entity.[43] It further considers the composition of types of damage awards (price erosion, lost profits or reasonable royalty) from 1995 to 2001 and 2002 to 2008, respectively, though it excludes NPE data from this analysis due to the fact that NPEs are generally not entitled to lost profit damages as they do not compete with the infringing entity.[44]

In addition, the 2009 PwC Study considers the success rates at trial of NPEs versus practicing entities, and further distinguishes between success on summary judgment versus at trial.[45] The study finds that NPEs were successful 29 percent of the time overall, compared to a 41 percent success rate for practicing patent-holders. Whereas NPEs were slightly more successful than practicing entities at trial, they were successful on summary judgment only 12 percent of the time compared with a 20 percent success rate for practicing patent-holders.[46]

5. Lemley & Shapiro Study

In their 2007 study of reasonable royalty awards, Lemley & Shapiro focus on the extent to which court-awarded royalty rates properly apply apportionment for multi-component technologies.[47] Their data set covered all cases reported in Westlaw from 1982 to mid-2005 that awarded a reasonable royalty.[48]

---

38 2009 PwC Study, *supra* note 97, at 4.
39 *Id.* at 20.
40 *Id.* at 6.
41 *Id.*
42 *Id.* at 7.
43 *Id.* at 10.
44 *Id.* at 11.
45 *Id.* at 12.
46 *Id.*
47 Lemley & Shapiro, *supra* note 50.
48 *Id.* at 2030.

Notably, their study was cited in the Senate Report for S. 515 for their finding that only 58 reasonable royalty awards were awarded from 1980 to 2005.[49] Lemley & Shapiro arrived at this count by including only "the subset of cases in which a court has written an opinion disclosing the royalty awarded."[50]

Lemley & Shapiro track the differences in royalty rates between different industries groups, and find variations in the average royalty rate awarded.[51] They conclude that "the reasonable-royalty rules do in fact accommodate component products but only to a limited extent."[52] They do not appear to consider the amounts of damage awards, what royalty base was used, or track the final outcomes after appeal.

6. Allison, Lemley & Walker Study

In their study, *Extreme Value Patents*, Allison, Lemley & Walker analyzed data on patent litigation from 2000 to 2007 provided by the Stanford IP Litigation Clearinghouse (the "Clearinghouse"). The authors identified from the Clearinghouse data every patent that had been litigated eight or more times between 2000 and 2007, a total of 106.[53] They further identified a random set of 106 once-litigated patents from the Clearinghouse data. The authors collected information about entity status, industry characteristics, and indicia of patent value (such as number of claims, forward-citations, and prior art citations).[54]

The relevant findings of the Allison, Lemley & Walker study are noted as follows:

- Litigation Rates by Industry:
  - Software and telecommunications patents are far more likely to be litigated, even over mechanical and chemical patents.
  - In particular, software-implemented business method patents comprise a large portion of the most-litigated patents group at 15%, compared to only 4% of once-litigated patents.
  - Mechanical and electronics patents make up the bulk of the once-litigated patent cases at 53% and 25%, respectively. Conversely, they are of only minor significance in the most-litigated patent set at 8% for mechanical and 1% for electronics.
- Patent Owners:
  - More than one-third of all litigated patents were sold to another owner after issue and before the lawsuit was filed.
  - Small entities that keep their patents rather than selling them tend to litigate less often than either large entities or purchasers of small entity patents.
  - Among the most-litigated patents, there are significantly more non-practicing entities than among the once-litigated patents.

---

49 *Id.*
50 *Id.* at 2031.
51 *Id.* at 2034-35.
52 *Id.* at 2035.
53 Allison, Lemley & Walker, *supra* note 99, at 4-5.
54 *Id.* at 5.

- Ownership of once-litigated patents is more diverse, with no one type of company or industry representing any significant percentage.

The authors provide an extensive discussion of their classification technique, and references to other relevant work.[55] Additionally, they provide categorization of the parties' entity type to assist in identifying indicia of strategic litigation practices.[56]

7. Opderbeck Study

A recent study by David Opderbeck conducts an independent empirical analysis of patent damage awards data.[57] Data was obtained from the Administrative Office of the Courts data files for civil cases decided from 2002 to 2007.[58] His analysis finds that "damage awards are widely and stochastically distributed, which suggests that most cases are being adjudicated according to their facts rather than according to some predisposition towards large awards."[59]

Opderbeck analyzes the distribution of patent infringement damage awards, finding a mean of $4.3M, median of $0.8M, standard deviation of $9.8M, and skewness of 3.97.[60] On this basis, Opderbeck concludes that "the range of awards varied widely . . . [suggesting] a lack of any pattern in the awards."[61] He further calculates the correlations between size of award and field of art. He finds "possibly significant" correlations with field of art of 0.36 (awards >= $500k), 0.54 (awards >= $1M), and 0.63 (awards >= $10M), but cautions that the sample sizes of the upper award tiers were small.[62]

Notably, Opderbeck further studies the correlation between size of award and type of remedy (lost profits or reasonable royalty). He finds correlations of 0.12 (awards >= $500k), 0.01 (awards >= $1M), and 0.52 (awards >= $10M).[63] From this, Opderbeck concludes that the sample reveals "no overriding patterns to the awards, except for some varying degrees of correlation between the size of award and the field of art or type of remedy."[64]

Opderbeck concludes that "the manner in which courts calculate reasonable royalty rates does not fundamentally cause any holdup and royalty stacking problems."[65] Instead, he suggests that "some facially shocking but mostly innocuous data are being used as the point of a much longer spear, which aims to redefine what kind of right a 'patent' represents."[66]

---

55 *See id.* at 6-11.

56 *See id.* at 12-14, *citing* Mark A. Lemley & Nathan Myhrvold, *The Complex Ecology of Patent Plaintiffs* (working paper 2009).

57 Opderbeck, *supra* note 100.

58 *Id.* at 145.

59 *Id.* at 130.

60 *Id.* at 146.

61 *Id.*

62 *Id.* at 148.

63 *Id.*

64 *Id.* at 149.

65 *Id.*

66 *Id.*

C. *Recent Patent Damages Case Law*

Several recent opinions, most issued in the 18 months prior to this article, have taken strides to reshape patent damages law and redress certain prevalent concerns. In particular, these cases have arguably heightened the standards for establishing reasonable royalty damages and instituted standards for more exacting scrutiny of jury verdicts.[67]

1. *Lucent v. Gateway*

In *Lucent v. Gateway*, a Federal Circuit panel vacated the jury's reasonable royalty award of $358 million for a minor component of Microsoft Office that was found to be infringing plaintiff's patent.[68] The issue on appeal was "whether substantial evidence supports the jury's implicit finding that Microsoft would have agreed to, at the time of the hypothetical negotiation, a lump-sum, paid-in-full royalty of about $358 million."[69]

The Court began by enumerating each of the *Georgia-Pacific* factors at issue and assessing the testimony and documentary evidence pertaining to each.[70] Principally relevant was the first factor, the "established royalty" for licensing the patents in suit. There, eight licenses that were accepted into evidence and used by the jury at trial were rejected as lacking "sufficient relevance"[71] to support the verdict. The jury had awarded a lump-sum royalty amount, but four of the licenses were based on running royalties and therefore were not comparable.[72] The other four licenses provided for lump-sum royalties but included additional material and arose under different circumstances than the hypothetical negotiation assumed, and therefore were not "sufficiently comparable."[73]

Accordingly, the Court reversed and remanded, having reached the "unmistakable conclusion that the jury's damages award is not supported by substantial evidence, but is based

---

67 *See* Lucent Techs., Inc. v. Gateway, Inc., 580 F.3d 1301, 1301 (Fed. Cir. 2009); ResQNet.com, Inc. v. Lansa, Inc., 594 F.3d 860 (Fed. Cir. 2010). Two district court opinions authored by Chief Judge Rader of the Federal Circuit, sitting by designation, also reflect this view. *See* Cornell Univ. v. Hewlett-Packard Co., 609 F. Supp. 2d 279 (N.D.N.Y. 2009) (C.J. Rader sitting by designation); IP Innovation LLC v. Red Hat Inc., No. 2:07-CV-447 (RRR), 2010 WL 986620 (E.D. Tex. Mar. 2, 2010) (C.J. Rader sitting by designation). Another recent Federal Circuit opinion reiterated the principles articulated in *Lucent* and *ResQNet* in reversing the district court's denial of defendant's F.R.C.P. 59(a) motion for a new trial on grounds that the damages awarded by the jury were "'clearly not supported by the evidence' and 'based only on speculation or guesswork.'" WordTech Sys., Inc. v. Integrated Network Solutions, Inc., 609 F.3d 1308, 1319 (Fed. Cir. 2010), (quoting Del Monte Dunes at Monterey, Ltd. v. City of Monterey, 95 F.3d 1422, 1435 (9th Cir. 1996)). And, in the first week of 2011, the Federal Circuit further supported this line of cases with its decision in *Uniloc USA, Inc. v. Microsoft Corp.*, __F.3d__, 2011 WL 9738 at *43 (Fed. Cir. Jan. 4, 2011).

68 *Lucent*, 580 F.3d at 1337 ("The only reasonable conclusion that can be drawn from this evidence is that the infringing use of Outlook's date-picker feature is a minor aspect of a much larger software program and that the portion of the profit that can be credited to the infringing use of the date-picker tool is exceedingly small.").

69 *Id.* at 1309.

70 Georgia-Pacific Corp. v. US Plywood Corp., 318 F. Supp. 1116, 1120 (S.D.N.Y.1970).

71 *Wordtech*, 609 F.3d at 1309.

72 *Lucent*, 580 F.3d at 1329-30.

73 *Id.* at 1328-29.

mainly on speculation or guesswork."[74] Subsequent decisions have followed this mode of careful analysis of the "sufficiency" of evidence of prior licenses.[75]

Notably, in closing, the Court also stated that to the extent the jury has applied the EMVR, this would have constituted legal error.[76] Though writing in dicta, the Court went to lengths to explore a long history of EMVR precedent, dating back to *Garretson v. Clark* in 1884.[77] The Court stressed the necessity for the plaintiff to establish applicability of the EMVR by demonstrating that that the patented feature constitute the "basis for customer demand."[78] Subsequent cases cite *Lucent* for these EMVR principles.[79]

2. *ResQNet v. Lansa*

In *ResQNet v. Lansa*, a Federal Circuit panel vacated a bench damages award of $506,305 for infringement of a patent directed to a computer terminal emulation algorithm, which award was calculated by applying a hypothetical 12.5% royalty rate to the defendant's revenues from sales of the infringing software.[80] At the outset, the Court emphasized that the fact-finder "must carefully tie proof of damages to the claimed invention's footprint in the market place,"[81] and cited its precedent for the rule that "[t]o prevent the hypothetical [negotiation] from lapsing into pure speculation, this court requires sound economic proof of the nature of the market and likely outcomes with infringement factored out of the economic picture."[82]

In arriving at a 12.5% reasonable royalty rate, the plaintiff's expert used average royalty rates from two sets of prior licenses to the patents in suit and related technology. One set of prior licenses related to re-branding and re-bundling licenses which "furnished finished software products and source code, as well as services,"[83] was rejected because the plaintiff had not shown that these licenses "embody or use the claimed technology" claimed by the patents in suit.[84] The other set were "straight" licenses to the patents in suit, which arose as settlements of prior litigation brought by the patentee.[85] The Court acknowledged that the settlement licenses could be admissible, but cautioned that even these must be scrutinized because settlement royalty rates may be too high (for example, "license fees negotiated in the face of a threat of high litigation costs may be strongly influenced by a desire to avoid full litigation"[86]) or too low (for example, "widespread infringement [could] artificially depress . . . past licenses"[87]), compared to

---

74 *Id.* at 1117.
75 *Wordtech*, 609 F.3d at 1309.
76 *Lucent*, 580 F.3d at 1336.
77 111 US 120, 121 (1884).
78 *Lucent*, 580 F.3d at 1336-37.
79 *See*, *e.g.*, Uniloc USA, Inc. v. Microsoft Corp., __F.3d__, 2011 WL 9738 at *51 (Fed. Cir. Jan. 4, 2011).
80 ResQNet.com, Inc. v. Lansa, Inc., 594 F.3d 869 (Fed. Cir. 2010).
81 *Id.*
82 *Id.* (*quoting* Grain Processing Corp. v. Am. Maize-Prods. Co., 185 F.3d 1341, 1350 (Fed. Cir. 1999)).
83 *ResQNet.com*, 594 F.3d at 870.
84 *Id.* at 871.
85 *Id.*
86 *Id.* at 872 (*quoting* Hanson v. Alpine Valley Ski Area, Inc., 718 F.2d 1075, 1078-79 (Fed. Cir. 1983)).
87 *Id.* at 872 (*citing* Nickson Indus., Inc. v. Rol Mfg. Co., 847 F.2d 795, 798 (Fed. Cir. 1988)).

what parties in an *ex ante* hypothetical negotiation would reach. Rather, the Court stressed use of prior licenses under *Georgia-Pacific* factor 1 must account for the "technological and economic differences" between the licenses and the assumptions underlying the hypothetical negotiation.[88]

3. *WordTech Systems v. Integrated Networks*

In *WordTech Systems v. Integrated Networks*, a Federal Circuit panel reversed the district court's denial of defendant's motion for a new trial in light of a jury damages award of $250,000 for infringement of a patent directed to a device for copying video files from computer memory to multiple discs.[89] The jury award calculated damages as a lump sum royalty (as opposed to a running royalty on sales),[90] based on evidence of thirteen past licenses to the patents in suit. Notably, the Court reiterated the lessons of its *Lucent* and *ResQNet* precedent that when using past licenses to calculate a reasonable royalty damages award, the licenses in the record must be "sufficiently comparable" on the basis of the circumstances and technology involved in each[91] and the comparison "must account for the technological and economic differences" to the present case.[92] Turning to the licenses in the record, the Court scrutinized each, finding that the amounts agreed to therein were substantially lower than the royalty amount (with respect to the licenses involving a lump sum royalty)[93] or effective royalty rate (with respect to the licenses involving a running royalty)[94] awarded by the jury.

4. *Cornell University v. Hewlett-Packard*

In *Cornell University v. Hewlett-Packard*, Federal Circuit Chief Judge Rader, sitting by designation, granted the defendant's motion for judgment of a matter of law, and in the alternative offered the plaintiff remittitur, in a damages amount of $58,494,282.[95] Although the final award was still substantial, it was less than one third of the jury verdict of $184,044,048.[96] Notably, Judge Rader's opinion did not address the royalty rate at all, which was an uncontested and minimal 0.8%, and focused solely on the issue of apportionment as applied to Hewlett-Packard's sales of CPU brick products containing, as a relatively small though functionally advantageous and "important component"[97] thereof, an instruction-issuing mechanism that infringed Cornell's patent.

Chief Judge Rader first articulated the requirements for applying the EMVR in a reasonable royalty analysis.[98] Further, Chief Judge Rader explained that sufficient evidentiary proof of the applicability of the EMVR is a "demand curve [or] market evidence indicating that

---

88 *Id.* at 873.
89 *WordTech Sys., Inc. v. Integrated Network Solutions, Inc,*, 609 F.3d 1308, 1309 (Fed. Cir. 2010).
90 *Id.* at 1310.
91 *Id.*, *quoting Lucent Techs, Inc. v. Gateway, Inc.*, 580 F.3d 1301, 1328-29 (Fed. Cir. 2009).
92 *Id.*, *quoting ResQNet*, 594 F.3d at 873.
93 *Id.*
94 *Id.* at 22.
95 Cornell Univ. v. Hewlett-Packard Co., 609 F. Supp. 2d 279, 293 (N.D.N.Y. 2009).
96 *Id.* at 282.
97 *Id.* at 285.
98 *Id.* at 286-87.

[the patented] invention drove demand for [infringer's products]."[99] Requiring such economic evidence of market demand, Chief Judge Rader rejected the plaintiff's expert's methodology of selecting the revenue base as sales of the CPU brick without showing the connection to consumer demand for the infringed component thereof.[100] The Court rejected this evidence, holding that "manufactured revenues cannot . . . sustain expansion of the [EMVR] beyond some credible economic indicators."[101]

Another recent district court case authored by Chief Judge Rader sitting by designation took a similar approach, holding that the plaintiff "must show some plausible economic connection between the invented feature and the accused operating systems before using the market value of the entire product as the royalty base."[102]

5. *Uniloc USA, Inc. v. Microsoft Corp.*

Most recently, the Federal Circuit issued its opinion in *Uniloc USA, Inc. v. Microsoft Corp.*, in which it rejected the long-standing "25% Rule of Thumb" for establishing a starting point for a reasonable royalty calculation.[103] The Rule was a common methodology used by plaintiffs' damages experts, whereby an initial royalty rate of 25% was assumed and case-specific factors were then applied to vary from that rate to arrive at a final number. In its opinion, the Court noted that while it had never squarely addressed admissibility of the Rule, the Federal Circuit has "passively tolerated its use where its acceptability has not been the focus of the case."[104] In premising its holding in the *Daubert* standard for expert evidence, the Court held the Rule to be inadmissible "because it fails to tie a reasonable royalty base to the facts of the case at issue."[105]

Explicitly, the Federal Circuit heavily relied and expanded on its precedent in *Lucent*, *ResQNet* and *Wordtech* in reaching its decision. The Court cited its precedent for the principle that "a patentee could not rely on license agreements that were 'radically different from the hypothetical agreement under consideration' to determine a reasonable royalty."[106] The Court emphasized that the "meaning of these cases is clear: there must be a basis in fact to associate the royalty rates used in prior licenses to the particular hypothetical negotiation at issue in the case."[107] Because the 25% Rule is "an abstract and largely theoretical construct [that] fails to satisfy this fundamental requirement," it was inadmissible as a tool for determining damages.[108]

Notably, the Court also harkened back to recent decisions regarding the EMVR in the second part of its opinion. Addressing the issue of application of the EMVR, the Court

---

99 *Id.* at 288.
100 *Id.* at 285.
101 *Id.* at 288.
102 IP Innovation LLC v. Red Hat, Inc., No. 2:07-cv-447(RRR), 2010 WL 986620 at *3 (E.D. Tex. Mar. 2, 2010).
103 Uniloc USA, Inc. v. Microsoft Corp., __F.3d__, 2011 WL 9738 (Fed. Cir. Jan. 4, 2011).
104 *Id.* at *39.
105 *Id.* at *41.
106 *Id.* at *43.
107 *Id.* at *45.
108 *Id.*

cautioned against the "danger of admitting consideration of the entire market value of the accused [product] where the patented component does not create the basis for customer demand."[109]

D. Implications of Recent Patent Damages Case Law and Questions for Study

In sum, *Uniloc* and its predecessors appear to strike a new course regarding damage awards and their methodologies and evidentiary foundations. It remains to be seen whether this shift will be substantive as well as rhetorical in the long term. Currently, the broader impact of these decisions, both on appellate review of patent infringement damage awards and on initial admissibility decisions at trial, is not yet empirically observable. It is possible that a central cause of excessive patent damages, to the extent they existed, has been corrected by these decisions. Or, these cases may have no long-term impact.

It is clear, however, that the Federal Circuit bench is taking an active role in reviewing patent damage awards and is seeking to clarify the rules for their determination at trial. To the extent problems with damages behave idiosyncratically, case-specific correction may be the most effective remedy.[110] Or, the opposite might be true. Since statutory changes operate differently than the organic evolution of case-by-case precedent, legislative patent reform might be more effective at fixing current problems in patent damage awards. Which fork should be taken depends on what, in fact, these problems are. That is, the nature of the appropriate remedy depends on the diagnosis of the problem.

More precisely, if excessive patent damages are found to behave idiosyncratically, then case-by-case correction of such individual errors and establishment of precedent to prevent their recurrence under analogous circumstances may be the best approach. However, if excessive damages are a systematic problem, legislative changes that would categorically impact all patent cases may be more effective. Or, as a third alternative, if excessive damages are systematically found across a subset of patent awards with identifiable characteristics, legislative or judicial approaches (or a combination of both) may be appropriate to target the problem. In this third case, identifying the characteristics in question will be key.

---

109 *Id.* at *51.

110 Burk and Lemley make this argument in their excellent book from 2009, *The Patent Crisis and How the Courts Can Solve It*. See Dan L. Burk and Mark A. Lemley, THE PATENT CRISIS AND HOW THE COURTS CAN SOLVE IT (2009).

II. Data

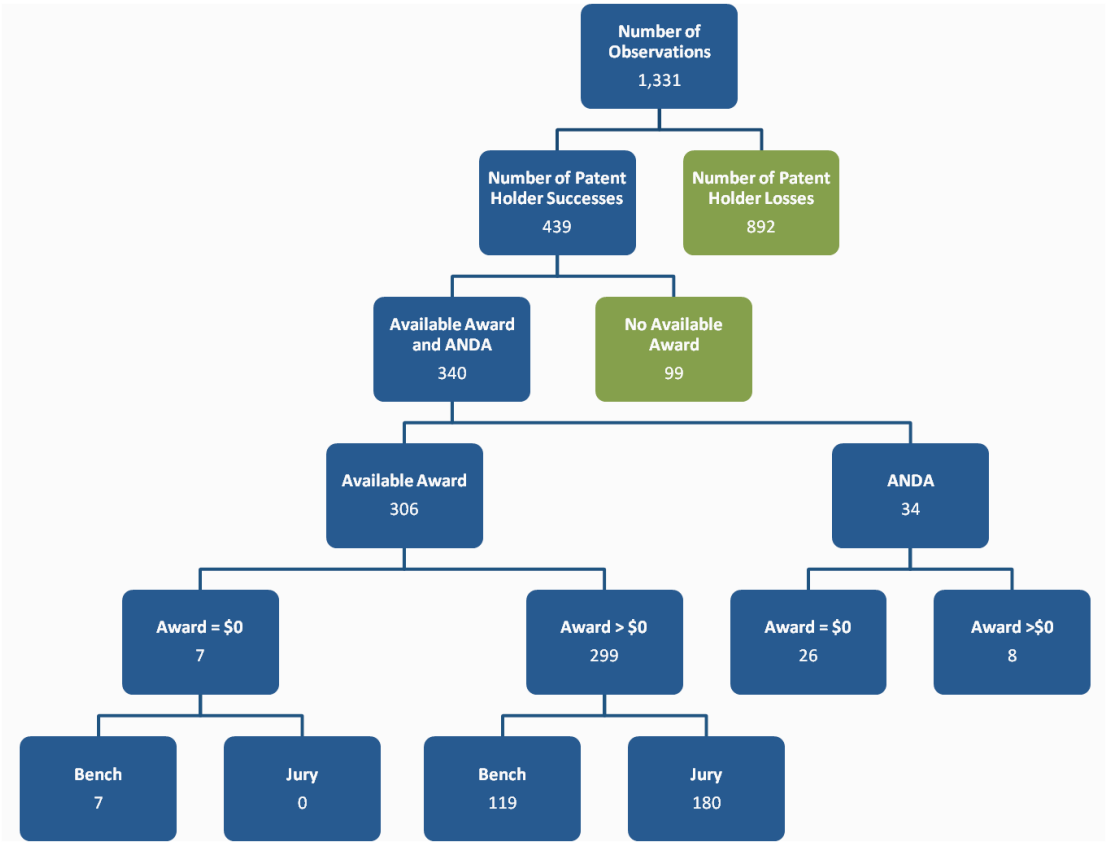
This study builds a comprehensive dataset of patent awards and attempts, to the extent possible given available information, to systematically characterize the distribution of damage awards. In addition, we construct a series of variables from a variety of sources that are subsequently used to predict and explain the size of awards in the dataset. This section of the paper discusses the dataset construction and provides a first glimpse of the information we have analyzed. We emphasize some interesting patterns in the raw data in this section, before presenting regression analyses in the next section.

A. Dataset

To start, our analysis requires comprehensive information about damage awards in litigated patent cases. As part of its intellectual property (IP) dispute analysis practice, which provides IP litigation and valuation services, PricewaterhouseCoopers LLP (PwC) has collected an extensive database on the complete set of patent case final rulings and damage awards as reported by Westlaw. Information in the PwC database includes party names, the industry of the potential infringer, whether the patent holder is a non-practicing entity, the presiding court at the time of the decision, the deciding body (bench or jury), the year of decision, the time to trial, and the associated damage awards with their component parts (where available). PwC updates its dataset every year and uses it to issue an annual report on statistics and trends in patent litigation and damages.[111] PwC licensed to us the proprietary dataset underlying their reports for the years 1995 through 2008 to start the process of building the dataset for this study. We carefully investigated each of the cases identified in PwC's original database to determine the nature of the intellectual property at issue and to verify that damage awards pertaining to the same litigated case were appropriately combined. After making a series of data cleaning changes, this process yielded a final case information database that is summarized in Figure 1.

[111] The most recent PwC studies are available at: http://www.pwc.com/us/en/forensic-services/publications/patent-litigation-study.jhtml. The PwC annual reports were often cited in the patent reform debates that preceded the passage of the America Invests Act.

FIGURE 1
Description of the Final Case Information Database
1995 – 2008



A total of 1,331 cases were identified, of which the trial court ruled there was infringement in 439. Among these, courts awarded damages in 340 cases – with post-judgment settlement by the parties being the most common reason no award data was found. These 340 cases represent the set of observations examined in this analysis, with the identified total damages award level representing the main dependent variable of interest.[112] The level of some of these awards may well have changed on post-trial review and appeal; however, attention is focused only on the initial damage awards granted at the district court level.[113] In other words,

---

112 The 340 cases include those involving Abbreviated New Drug Applications (ANDAs) where lost profits and reasonable royalties are not available remedies. To avoid losing these cases in the regression analysis they are coded as having $0 award (if there were no costs awarded). Because some total damages amounts include costs that cannot be separated out, all total awards include costs and attorneys fees, where available. Further, seven non-ANDA cases have a true award of $0. In these cases, the trier-of-fact determined that the patent holders did not bear their burden of proof on damages.

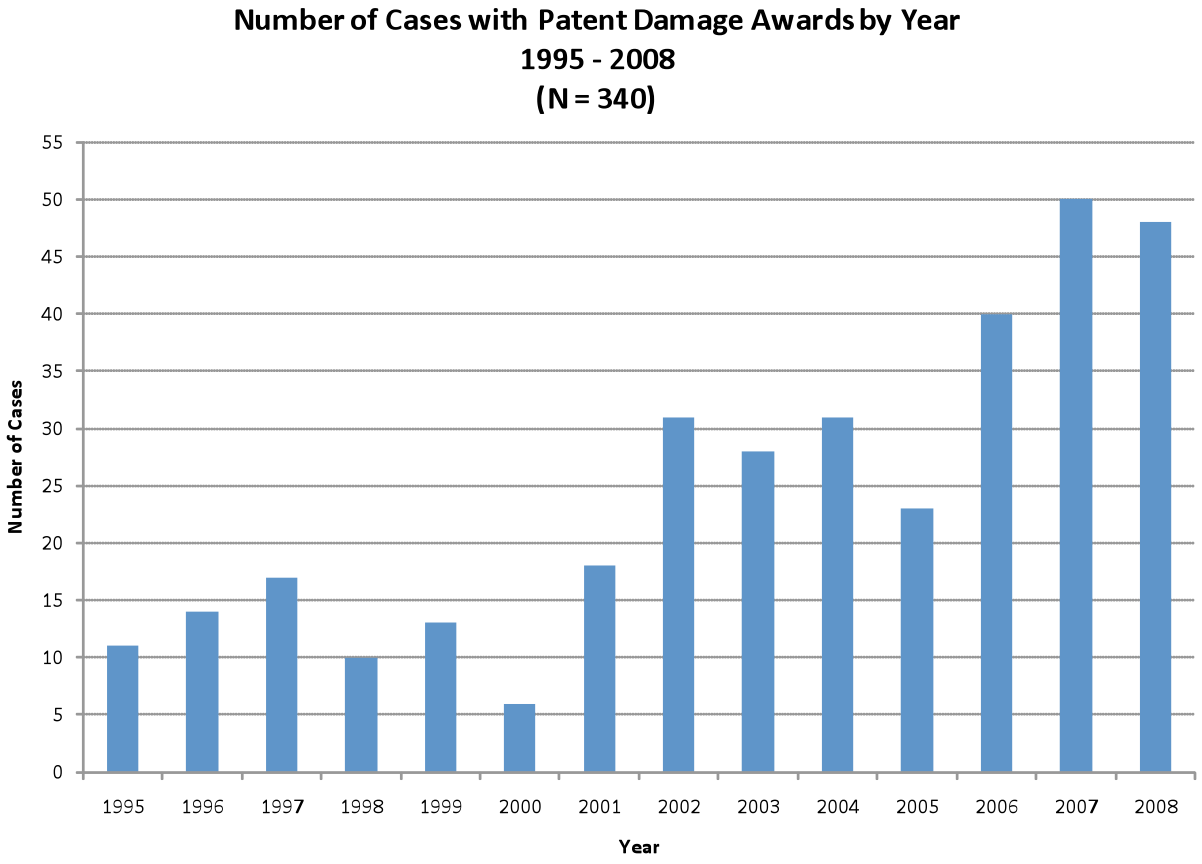
113 To be clear, we define awards based on the trier of fact in the case. For cases decided by a jury, the base amounts are those awarded in the jury verdict. For cases decided by a judge, the base amounts are the those in the final judgments. Base awards are for direct infringement only (including price erosion and convoyed sales where awarded). They do not include appeals or, in the case of jury awards, remittiturs by the bench. Where available, associated interest and enhanced damages for willfulness are added to the base amounts to arrive at the total award.

the damage awards in our dataset may have been changed during the appeals process, but these changes are not reflected in our current analysis.[114] To compare across years, we used the Consumer Price Index to translate damage awards levels from their nominal amounts into 2008 dollars.

B. Characteristics of the Award Distribution

Figure 2 displays the count of observations in the datatset by year of decision, from 1995 through 2008.

FIGURE 2



This graphic representation underlines the fact that on a year-by-year basis, the number of patent damage awards granted is quite small. As a consequence – and particularly since one or two large awards can skew these distributions substantially – one should be careful to not attribute too much significance to differences in observed damages from year to year.[115] In fact,

---

114 Future analysis may study the changes in awards due to the appeals process.

115 Another reason for caution in making year-to-year comparisons is because of the E-Government Act of 2002 (Pub.L. 107-347, 116 Stat. 2899, 44 U.S.C. § 101, H.R. 2458/S. 803) which applied to the federal judiciary and mandated public electronic access to all written court case opinions. This Act could account *(cont'd)*

wWhen controlling for the year of the decision in some of the regressions below it can be shown that an independent time trend has very littleno power in explainingis negatively correlated with damages award amounts.[116]

To facilitate comparison with previous studies, annual summaries of the distributions of awards in the dataset are presented. Table 12 provides a more complete picture of these distributions, by including the quartiles as well as medians.[117] Taking 2004 as an example, after adjusting the awards to 2008 dollars, the lowest award that year was $40,000 and the highest award that year was $175.1 million. In between those amounts though, 25% of the awards were under $540,000, 50% of the awards were under $4.3 million, and 75% of the awards were under $29.0 million. The other annual distributions behave in similar fashion.

TABLE 1
Distribution of Patent Damage Awards by Year ($ in millions, 2008)
1995 – 2008
(N = 306)

Year	Minimum	First Quartile	Median	Third Quartile	Maximum
1995	$0.03	$1.38	$5.07	$16.32	$87.52
1996	$0.02	$0.37	$3.57	$22.68	$130.36
1997	$0.30	$1.55	$7.70	$24.03	$97.59
1998	$0.01	$2.18	$3.81	$10.63	$225.87
1999	$0.28	$1.95	$7.35	$20.97	$125.35
2000	$0.48	$0.61	$3.02	$6.59	$16.54
2001	$0.00	$0.08	$1.58	$16.91	$94.87
2002	$0.00	$0.61	$5.15	$30.77	$117.41
2003	$0.08	$0.70	$10.41	$19.93	$609.17
2004	$0.04	$0.54	$4.27	$28.99	$175.09
2005	$0.00	$1.92	$8.23	$26.92	$141.14
2006	$0.01	$0.44	$2.94	$32.22	$327.76
2007	$0.00	$0.14	$1.11	$18.12	$1,597.11
2008	$0.00	$0.66	$2.88	$27.18	$1,223.88

(cont'd from previous page)

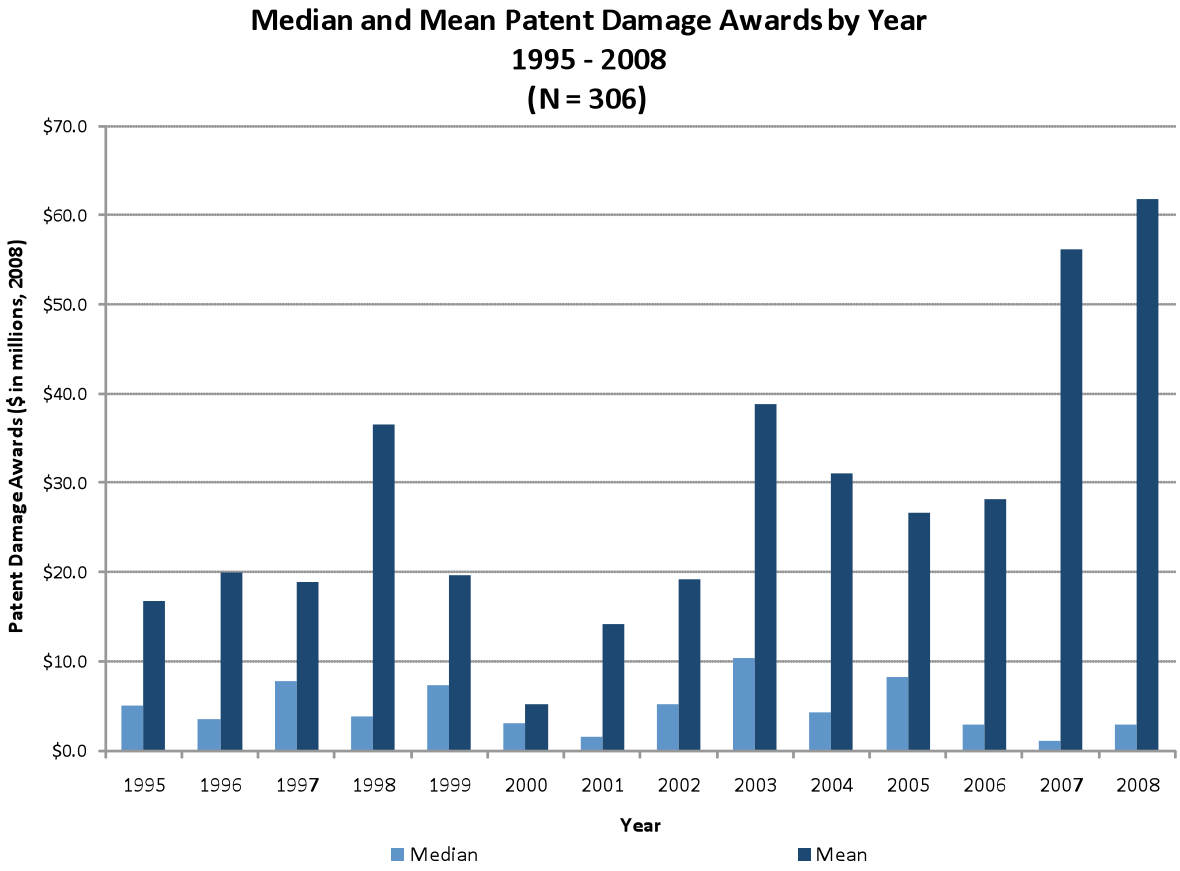
for the increase in cases starting in 2002 and going through 2008 as more courts implemented the requirements in the Act.

[116] Furthermore, the small number of patent infringement cases in which damages are awarded may give reason to question the hyperbolic claims by some that patent litigation damages have significant deleterious effects on research and development activities in the United States.

[117] Since patent cases involving Abbreviated New Drug Applications (ANDAs) are structured differently from standard patent infringement cases in terms of damages, those cases were removed from Table 2 as well as from Figures 3 – 6 for descriptive purposes. The total number of cases without ANDA cases is 306 rather than 340.

Figure 3 shows the differences in the median and average damages awards by year.

FIGURE 3



Although there is an underlying stability of the median over time, the increasing skewness of the awards data is evident from Table 2 and Figure 3 – for example, when they occur, outliers generate large differences between the average and the median award levels in particular years. Taken together, Table 2 and Figure 3 also demonstrate an underlying stability of the distribution over time. This lack of annual variation motivates a description of the characteristics of the *entire* distribution of awards over the whole time period for which data is available.

A straightforward graphical presentation of the entire awards distribution is shown in Figure 4.

FIGURE 4

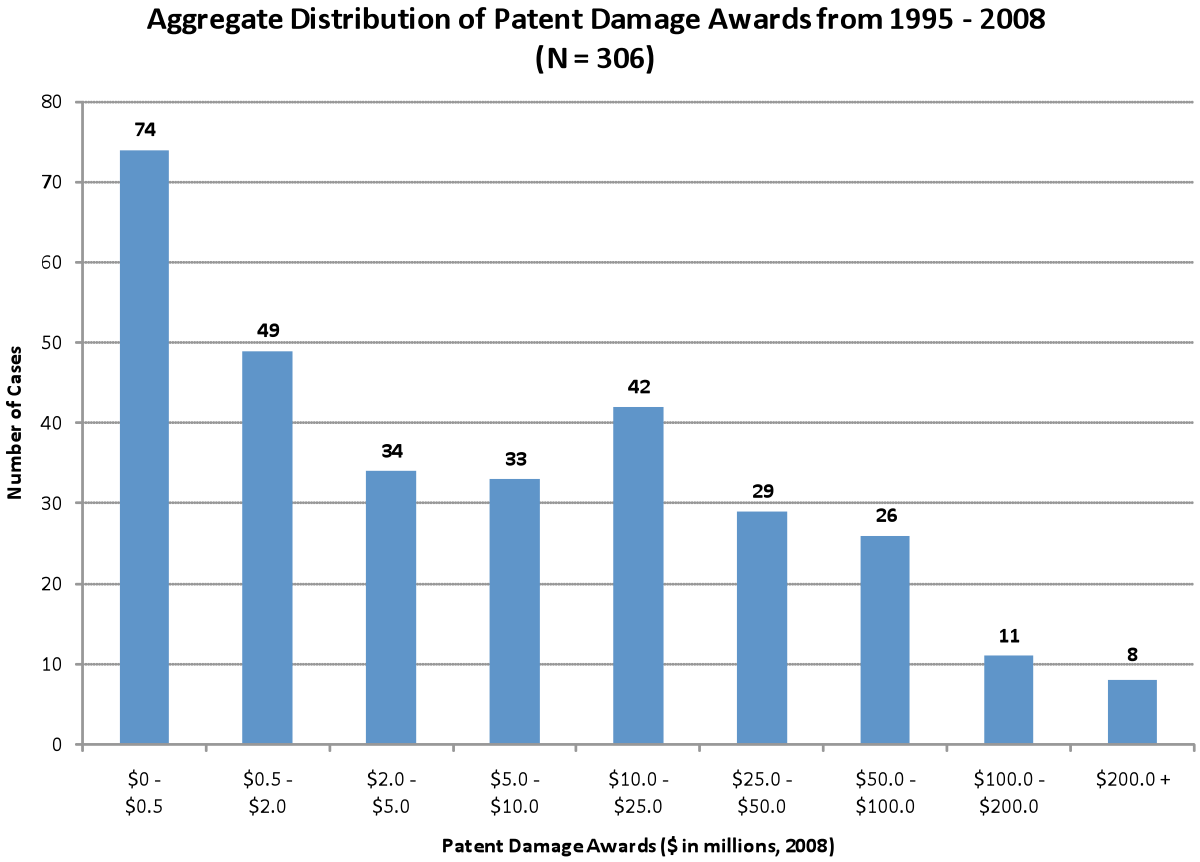


Figure 4 is a histogram of awards, broken down into increasing award-level categories. Across the dataset, 74 of the cases have damage awards of less than $500,000, representing 24.2 percent of all cases during the time period. Reading from left to right in the figure, 49 cases have award values between $500,000 and $2 million; 34 between $2 and $5 million; 33 between $5 and $10 million, 42 between $10 and $25 million, 29 between $25 and $50 million, 26 between $50 and $100 million and 11 between $100 and $200 million. Of particular note in Figure 4 is the very last bar on the right, representing damage awards of over $200 million. A total of eight cases fall into this highest category of damage awards, which represents 2.4 percent of the number of all awards during the 1995 through 2008 period.

It is not surprising that these damage awards in the upper tail of the distribution would attract so much attention. As compared to the overall distribution, they are quite large. Indeed, we find that together, these eight cases represent 47.6 percent of the collective damages in *all* the non-ANDA cases from 1995 until 2008. These raw data suggest that focusing on these very large values may obscure the true nature of the overall distribution of the damage awards. In contrast to the suggestion put forward by policy makers prior to the enactment of the America

Invests Act,[118] our systematic analysis of the entire distribution reveals that the largest awards are not "the tip of the iceberg" of excessive patent damage awards. Instead, these very large awards appear to be true outliers, as compared to the rest of the distribution.

While more details about the determination of awards will be discussed in the regression analysis described below, a descriptive analysis of the underlying distribution of damage awards is revealing about concerns regarding the unpredictability of patent damage awards. Cutting the data several ways shows that the distribution exhibits a great deal of skewness; a very small number of very large damage awards are not representative of what has happened across all cases. This may yet be another example of the behavioral bias that occurs when individuals "overreact" to the very low probability, but very bad outcomes.[119] Nonetheless, it is notable that such awards are indeed very large as compared with the rest of the distribution.

C. Explanatory Variables

To complement the damage awards information, we also assembled various series of data that could potentially explain the level of damages in each case. All the explanatory variables used are summarized in Table 2 and can be divided into three separate categories.[120] The first category is information derived from the record in each individual case, with key factors such as whether the case was decided by a judge or a jury and whether a lost profit or a reasonable royalty damages theory was utilized in determining the level of the award, if available.

The second category of variables represents information about the litigants in each case. This includes the identity of both the plaintiff and the defendant in each case—*i.e.*, if it is an individual, a firm, a government entity or a nonprofit organization. The corporate litigants are further broken down into various industry categories and by firm size.

The third category of variables draws on the economic literature of patent value mentioned above. These data include publicly available information on various characteristics of patents, including information about their assignees, number of claims, and counts of their citations in subsequent patents. Economists have argued that patents embodying more substantial intellectual property often have more claims and are cited more often by later patents.[121] By including number of claims or appending citation information to the data for each case, it can be determined whether a particular measure of a patent's value is associated with the court's determination of infringement award levels.

All of the case identification and variable coding are limited to the information that could be found in Westlaw, Lexis, PACER, and the NBER patent database, in addition to information on websites like Google, Manta, Hoover's Online, Fortune, and EDGAR (for company SEC filings).[122]

---

118 2007 Senate Report, *supra* note 79, at 12.

119 See, for example, Cass R. Sunstein and Richard Zeckhauser, "Overreaction to Fearsome Risks," HKS Faculty Research Working Paper Series, December 2008.

120 For a list of variables, *see* Appendix 1.

121 *See* Allison et al., *Valuable Patents*, *supra* note 184.

122 The databases can be found at the following websites – Westlaw: https://lawschool.westlaw.com; Lexis: http://www.lexisnexis.com/lawschool; PACER: http://www.pacer.gov; NBER patent database: *(cont'd)*

TABLE 2
Summary of Variables

Variable Groups	Description	Sources
Category 1: Case Information		
Identifiers	Variables including a unique ID assigned by the authors, the docket number of the case, and the full names of the first listed plaintiff and defendant in the case.	PwC database, Google, Westlaw, and PACER
Dates	Variables including the year of the original award in district court, date the complaint for case was filed, the earliest start date of trial on validity, infringement, or damages, and the number of days between the trial start date and the complaint date.	PwC database, Google, Westlaw, and PACER
Location	Variables including where the case was litigated, including state, circuit, and court.	PwC database, Google, Westlaw, and PACER
Other Case Information	Variables determining if the case contained a summary judgment for the patent holder on validity and/or infringement, if the case involved an invalidated patent-at-issue, and if the patent holder was successful in its patent claims.	PwC database, Google, Westlaw, and PACER
Damage Awards	If the patent holder was successful, variables for the total award amount, lost profits, reasonable royalties, prejudgment interest, enhanced damages, price erosion damages, and other damages. Also included are whether or not the case settled before damages were awarded, whether or not the case resulted in only an injunction, and whether or not the case was an ANDA filing.	PwC database, Google, Westlaw, and PACER
Category 2: Litigant Information		
General Assignee	Includes number of patent assignees associated with the patents-at-issue in the case, the names of the assignees, if one of the assignee(s) is the first named plaintiff or defendant in the case (can be both), if the plaintiff name listed is an assignee (patent holder), and if the patent holder markets or manufactures its technology covered by the patent.	PwC database, Google, Westlaw, PACER, and NBER patent database
NBER Assignee	Dummy variables from the 2002 NBER database which coded the Assignee(s) as "Unassigned," "US, Non-Government," Non-US, Non-Government,", "US, Individual," "Non-US, Individual," "US Government," or "Non-US, Government."	NBER patent database
Assignee Identifiers	Includes the variables determining whether or not the first named plaintiff or defendant are an individual, private entity, public entity, university, part of the U.S. government, a domestic entity, foreign entity, part of the 2009 Fortune 500 list, part of the 2009 Fortune 1000 list, a subsidiary of a parent company.	EDGAR, Manta, Hoover's Online, Westlaw, and Fortune 1000
Assignee Parent Identifiers	Variables for the parent companies of the plaintiff or defendant listed if it was a subsidiary that include whether or not the parent company is a private entity, public entity, domestic entity, foreign entity, part of the 2009 Fortune 500 list, part of the 2009 Fortune 1000 list, if the first named plaintiff or defendant is owned by a joint venture (2 parents or more).	EDGAR, Manta, Hoover's Online, Westlaw, and Fortune 1000
SIC Codes	Variables identifying the 2-, 3-, and 4- digit SIC codes for the potential infringers.	NBER patent database, Google, and Westlaw
Category 3: Patent(s)-at-Issue Information		
General Patent	Variables identifying the number of patent(s) at issue in the case and their type as either utility, reissue, design, or application number.	NBER patent database, Google, and Westlaw
Patent Classification	Includes variables for all patents-at-issue such as application year calculated for minimum and maximum (minimums and maxima differ for cases with multiple patents-at-issue and are the same for cases with only one patent-at-issue); grant date year calculated for minimum and maximum; grant date calculated for minimum and maximum; age of the oldest and youngest patent-at-issue in a case calculated for minimum and maximum; number of claims calculated for minimum, maximum, average and total; number of forward citations through 2002 from the NBER 2002 data, calculated for minimum, maximum and average; number of forward citations through 2010 if the 2002 forward citations were not available, calculated for minimum, maximum and average; the IPC4 classification listed first on the patent; and the PTO main classification for each patent listed in the case.	NBER patent database, Google, and Westlaw

(cont'd from previous page)

http://elsa.berkeley.edu/~bhhall/patents.html and https://sites.google.com/site/patentdataproject/Home; Google: http://www.google.com; Manta: http://www.manta.com; Hoover's Online: http://www.hoovers.com; Fortune 1000: http://money.cnn.com/magazines/fortune/fortune500/2009/full_list/; and EDGAR: http://www.sec.gov/edgar.shtml.

III. EMPIRICAL ANALYSIS

A. Analysis of Overall Predictability

Using the dataset described in the section above, we first attempt to determine whether patent damage awards are predictable based on *ex ante* factors. Because our dataset does not contain the outcome of every patent case filed, we cannot yet create a model to predict the expected value of damages from the outset of a case. However, we can develop a model that explains damages conditional on the patent being found valid and infringed and the parties not settling. In future research, we plan to delve more deeply into the expected value of a filed patent case.

The regression analyses presented below attempt to determine how much of the variation in patent damage awards can be explained by the factors we assembled regarding the cases, litigants and patents-at issue. Using all 340 patent damage awards,[123] we ran several models to see which collection of factors could best be used to explain the variation in observed patent damages from 1995 through 2008. Because the dependent variable remains the same for most of the models, the R-squared goodness of fit measure can be used to compare the different models. The summary statistics from the models of best fit are outlined in Table 3.

TABLE 3
Summary of Models That Explain Patent Damage Awards[124]

Dependent Variable = Patent Damage Awards in 2008 $	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)
R-Squared	0.6399	0.7340	0.7403	0.7427	0.7561	0.7702	0.4457
Adjusted R-Squared	0.5368	0.6566	0.6621	0.6599	0.6618	0.6696	0.2030
F (k-1, N-k)	5.88 (75, 262)	15.15 (76, 261)	14.40 (78, 259)	20.44 (82, 255)	20.12 (94, 243)	19.50 (95, 217)	2.54 (95, 217)
Sample Size (N)	338	338	338	338	338	313	313
Standard Errors	Robust	Robust	Robust	Robust	Robust	Robust	Robust
Dependent Variable Type	Log	Log	Log	Log	Log	Log	Linear
Independent Variables	Base Controls	Model (1) + ANDA Dummy	Model (2) + Interactions	Model (3) + non-parametric total patents	Model (4) + Year Dummies	Model (5) + Avg. Forward Citations	Model (6)

Model (1) in Table 3 is our "naïve" model that contains almost all of the variables listed in Appendix 1 as controls. Because of the skewness inherent in the distribution of damages, the dependent variable is the log of damages in 2008 dollars. This transformation is necessary to normalize damages and allow for a better predicting model, as we will show. Also, to minimize multicollinearity that could artificially increase the R-squared goodness of fit measure, the

123 In order to compare all patent infringement cases, ANDA cases are included with $0 damages amounts in the regression. In most of the regressions fewer than 340 cases are used in the model due to missing data.

124 Full regression results are on file with the authors.

control variables were tested to ensure none were highly correlated with any other. For pairs of controls that were highly correlated, the one of the pair most correlated with the log of damages (the dependent variable) was retained. Robust standard errors were also used to mitigate any heteroscedasticity in the model.

This naïve model does quite well as it explains about 64 percent of the variation in the observed patent damage awards, as represented by an R-squared of 0.6399.[125] However, we thought it was possible to create an even better model by adding in or creating additional controls. In order to most effectively use the data to generate additional explanatory power, we conducted a variety of detailed manipulations on several of the variables. For example, we constructed interaction terms for certain key variables. As an illustration, the data contains information about who decided damages (judge or jury) in each case and the particular damages theory (lost profits or reasonable royalties) utilized. Based on these individual indicator variables, we created, for example, interaction variables for cases decided by juries using the reasonable royalty standard. We also considered nonlinear representations of some regressors. The remaining models in Table 3 show how each modification improved the overall predictability of patent damage awards.

Model (2) in Table 3 is the naïve model plus an additional control for whether the case was an ANDA case. Because ANDA cases generally have $0 awards, as a group they are different from standard patent infringement cases. Rather than drop these observations, we chose to control for them in Model (2). This addition immediately increased the explanatory power of the model as represented by its R-squared of 0.7340 (adjusted R-squared = 0.6566).

Acknowledging that juries having to decide reasonable royalty damages could influence the total amount of damages awarded, we added two interaction variables (juries x reasonable royalties and juries x lost profits) to create Model (3) in Table 3. These additions result in a minor improvement over Model (2), explaining 74 percent of the variation in damages (adjusted R-squared = 0.6621).

Model (4) in Table 3 uses Model (3) but replaces the single variable representing the number of patents-at-issue in the case with a non-parametric set of variables. This substitution suggests there may be diminishing marginal returns to each patent-at-issue in the case as to damages. A single variable suggests that each additional patent-at-issue contributes equally to the total damages awarded. The set of non-parametric variables, however, allows us to assess the explanatory power of groups of patents-at-issue on total damages. In Model (4), the set of non-parametric variables include dummies for cases with 1, 2, 3, 4, 5-10, and over 10 patents-at-issue. As with all sets of dummy variables, one is dropped for the regression, in this situation we drop the dummy for cases with only one patent-at-issue. Again, this model provides a minor improvement over Model (3) with an R-squared of 0.7427 (adjusted R-squared = 0.6599).

In Model (5) of Table 3, we replace a single variable for the year of the case decision with a set of dummy variables representing each year. This alternative means of accounting for the time trend of damages does improve the overall predictability of damage awards, although

[125] Even after taking into account the number of regressors in the model, the Adjusted R-squared still equals 0.5368.

none of the years are individually significant so the time trend itself does not have much predictive power. This model as a whole explains 76 percent of the variation in patent damage awards (adjusted R-squared = 0.6618).

Finally, Model (6) takes Model (5) and adds a variable tracking the average number of forward citations for the patents-at-issue in the case. Allison, et al. originally linked the number of forward citations to the likelihood of patent litigation[126] and forward citations as a proxy for the inherent economic value of patents appears to hold in this model as well. Model (6) explains about 77 percent of the variation in patent damages (adjusted R-squared = 0.6696). However, we note that forward citations, gathered from the NBER patent database, were not available for all cases and so had to drop 25 cases due to lack of data.

Remarkably, the statistical models that we constructed includes sets of regressors that explains between 64 and 77 percent of the variation in the observed patent damage awards. This result suggests that infringement damages are very predictable based on the dimensions represented by our data.[127] It is worth noting that the dependent variable in Models (1) –(6) reported in Table 3 is the log of damage awards. The skewness in the underlying damages data suggests this was a necessary transformation to determine a model of best fit since patent damages are not determined by a straight line (especially as they get larger). [128] Linear versions using the same regressors would have much less explanatory power. For example, Model (7) in Table 3 is simply Model (6) but with the straight patent damage amounts, i.e. without the log transformation. This model does far worse than any of the others (R-squared = 0. 4457 and Adjusted R-squared = 0.2030).

---

126 *See* Allison et al., *Valuable Patents*, *supra* note 184

127 These findings contrast with the suggestion in the Opderbeck study that there is no clear pattern to the observed damage awards. *See* Opderbeck, *supra* note 100, at 149.

128 It is not uncommon to use log transformations on the dependent variable in order to put the relationship between the dependent and independent variables into a linear form. The appropriateness of logging the dependent variable can also be determined by graphing the residuals of the model. Here the residuals are normally distributed, suggesting that our model is appropriate.

B. Key factors explaining patent damage awards

While the previous model focuses on the overall predictability of patent damage award levels based on observable factors, the relatively large number of regressors and the presence of interaction and higher-order terms complicates interpretation of individual explanatory factors. In this subsection, we present a streamlined version of the regression analysis, with regressors specifically chosen to assess various economic factors that may be associated with damage awards. In addition, we evaluate the role of certain litigation strategy and case-related variables that may also affect damage award levels.

The regression results are presented in Table 4. Again, the dependent variable of the regression is the natural logarithm of observed patent damage awards. Note that the number of observations in this dataset is somewhat smaller, as several cases needed to be dropped due to incomplete data for some of the important explanatory variables. Despite the much smaller number of explanatory variables (just ten) in this regression, the overall fit of the regression remains relatively strong.

The focus of this empirical exercise, however, is on the significance of the individual regressors. We start at the top of table 4 with four variables regarding the patents at issue in the decided cases. The number of patents varies by case (ranging from one to twenty-nine patents), and these results indicate that cases with more patents tend to have higher damage award values, all else equal. This factor had particularly high statistical significance, with a t-statistic of 4.99.

The next three explanatory variables capture features of the patents in each of the cases for which damages were awarded. Since there may be several patents associated with a given case, we included averages for each of these features calculated across the patents in that case. For example, based on the issue date of the patent and the time of the decision, we determined the age of each patent associated with the case and computed the average among all these patents. Again, the coefficient on average patent age is positive and statistically significant. Cases associated with more mature patents – perhaps those for which infringement would have generated a higher level of lost profits – do have a correspondingly higher level of damage award values.

The remaining two variables are meant to proxy for the inherent economic value associated with the patents-at-issue in the cases. For each case, we have computed the average claims made by the relevant patents; our hypothesis is that patents with more claims should cover more intellectual property. The resulting damage awards were indeed higher in cases where the patents had more claims, potentially reflecting a higher royalty rate or greater amount of lost profits related to more intellectual property. The significance of the intellectual property associated with patents is often captured by the number of times the patent is cited in other patents granted in the future. Our regression results support this interpretation as well, as damages are higher in cases where the average patent is cited more often in future patents. The regression coefficients on both the patent claims and forward citations variables are statistically significant.

TABLE 4
Significant Factors Influencing Damage Awards

Number of obs	240
F(10, 229)	15.710
Prob > F	0.000
R-squared	0.362
Root MSE	88629.000

Dependent = *Log of patent damage awards in 2008 dollars*	Coef.	Robust Std. Error	t	P>t	[95% Conf. Interval]	
Average Number of Patent Claims	0.00418	0.00169	2.47	0.014	0.00849	0.00751
Number of Patents	0.07319	0.01466	4.99	0.000	0.04431	0.10208
Average Number of Forward Citations	0.00526	0.00182	2.89	0.004	0.00168	0.00884
Average Age of Patent	0.00009	0.00004	2.31	0.022	0.00001	0.00016
Patent_manuf_mkt_tech = 1	0.18153	0.13329	1.36	0.175	0.08111	0.44417
Defendant is a Fortune 500 Company (or subsidiary)	0.25912	0.18626	1.39	0.166	0.10788	0.62613
Defendant is a Public Company (or subsidiary)	0.63925	0.13479	4.74	0.000	0.37367	0.90482
Dummy for Jury Trial	0.77575	0.15008	5.17	0.000	0.48003	1.07146
Year of Decision (time trend)	-0.05784	0.01557	-3.72	0.000	0.08851	0.02717
Time-to-Trial	0.00032	0.00008	4.06	0.000	0.00017	0.00048
Constant	120.59220	31.11397	3.88	0.000	59.28595	181.89850

The next set of reported coefficients is associated with the litigants involved in the cases. Unfortunately, we do not have specific information about the infringing activity that would allow us to directly measure lost profits or reasonable royalties on a case-by-case basis. We instead use variables associated with the size and revenue potential of defendants to proxy for the scope of what these damage values might be. We include dummy variables indicating cases where the defendant is a public company (as opposed to a private company, an individual or a government organization) and another dummy variable for those companies that are in the Fortune 500 (the 500 largest companies by revenue in the United States). Both of these dummy variables are positive, though only the public company proxy is statistically significant at traditional precision levels. Though these proxies are imperfect, these findings do provide some consistent evidence regarding revenue potential.

On the plaintiff side, we wanted to see whether there was a difference between litigants that were in business potential producing products based on their patents and so-called "non-practicing entities" (NPEs) who own patents and may assert rights even if they are not involved in associated commercial activity. Our proxy for this takes the form of a dummy variable (assembled by PwC) that equals one in cases where the patent holder manufactures and/or markets the technology associated with the patent. The regression results indicate that damage awards are higher in cases with such patent holders, though the estimated coefficient is not statistically significant. A positive result is not surprising, given that the nature of NPEs makes it difficult for them to assert a lost profits damages argument. Nonetheless, this result is worthy of further examination – in future work, we plan to investigate the role of NPEs in damage awards in greater depth.

The last set of variables in the regression focus on litigation-related factors, including case strategy choices that may be affected by litigants. We included a dummy variable for cases that were decided by juries; such cases were associated with significantly higher damage awards. Patent reform proponents cited case complexity and jury inexperience as contributors to "excessive" awards; our results do indicate higher awards in jury cases (though it is difficult to argue that the awards are "excessive" based on these results). We also include a time-to-trial variable that equals the number of days between the initial complaint and the date of the decision. While there are a variety of potential explanations for why the time to trial might be longer, we believe it may proxy for the complexity of cases – with more complex cases having potentially higher damages at stake.

Finally, we included the year of the decision in the regression to control for any time trend in the damage award amounts. Interestingly, the estimated time trend is significantly negative here, indicating that all else equal damage awards have been *decreasing* over time. To the extent that observed damage award values may have been increasing, the results suggest that this is more due to changes in the kinds of cases involved (as captured by our control variables) as opposed to any general independent trend toward greater awards. Indeed, to the extent there is an independent time trend it appears to be moving in the opposite direction.

IV. INTERPRETATIONS AND CONCLUSIONS

To be completed.

Appendix 1

List of Variables and Descriptions

Variable	Description	Source
Category 1: Case Information		
case_ID	Unique identifier for each case	Assigned
docket_number	The docket number associate with the case	PwC database, Google, Westlaw, and PACER
P_name_1	Full name of the first plaintiff listed on the case	PwC database, Google, Westlaw, and PACER
D_name_1	Full name of the first defendant listed on the case as reported in Westlaw	PwC database, Google, Westlaw, and PACER
P_pat_owner	The plaintiff is the patent holder	PwC database, Google, Westlaw, and PACER
year_of _decision	The year associated with the leading decisions entered by PwC	PwC database, Google, Westlaw, and PACER
complaint_date	Date the complaint for the case was filed	PwC database, Google, Westlaw, and PACER
trial_start_date	The earliest start date of a trial on validity, infringement, or damages	PwC database, Google, Westlaw, and PACER
time_to_trial	The number of days between *trial_start_date* and *complaint_date*	calculated
state	The state in which the court is located	PwC database, Google, Westlaw, and PACER
circuit	The circuit to which the court belongs, if a federal court; Additionally: 0= the U.S. Court of Federal Claims and 12 = D.C. Circuit; State courts are left blank	PwC database, Google, Westlaw, and PACER
court	The court in which the decision on patent holder success was made	PwC database, Google, Westlaw, and PACER
jury	The decision on damages made by a jury	PwC database, Google, Westlaw, and PACER
SJ_flag	The case contained a summary judgment for the patent holder on validity and/or infringement	PwC database, Google, Westlaw, and PACER
invalid_pat_flag	The case involved an invalidated patent-at-issue	PwC database, Google, Westlaw, and PACER
dmg_awd_flag	The patent holder was successful in it's patent claims; i.e. the patent was found valid and/or infringed	PwC database, Google, Westlaw, and PACER
dmg_awd_amt	The total dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages (if applicable)	PwC database, Google, Westlaw, and PACER

List of Variables and Descriptions (cont.)

Variable	Description	Source
	Category 1: Case Information (con't.)	
LP_flag	The patent holder was awarded lost profits	PwC database, Google, Westlaw, and PACER
LP_amt	The lost profits dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
RR_flag	The patent holder was awarded reasonable royalties	PwC database, Google, Westlaw, and PACER
RR_amt	The reasonable royalty dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
RR_rate	The percentage rate associated with the reasonable royalty dollar award granted to the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
RR_basis	The basis to which the reasonable royalty rate will be applied in the trial case before appeal of damages; =0 if the rate is not given in the decision or there is no rate; =1 if Sales; =2 if Profit	PwC database, Google, Westlaw, and PACER
PJI_flag	The patent holder was awarded prejudgment interest	PwC database, Google, Westlaw, and PACER
PJI_amt	The prejudgment interest dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
PJI_rate	The percentage rate associated with the prejudgment interest dollar award granted to the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
PJI_basis	The basis from which the prejudgment interest rate is derived in the trial case before appeal of damages; =0 if the rate is not given in the decision or there is no rate; =1 if Prime Interest Rate; =2 if Treasury Bills; =3 if Statutory Rate; =4 if Cost of Capital; =5 if given but Other	PwC database, Google, Westlaw, and PACER
enh_dmg_flag	The patent holder was awarded enhanced damages	PwC database, Google, Westlaw, and PACER
enh_dmg_amt	The enhanced damages dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
PE_flag	The patent holder was awarded price erosion damages	PwC database, Google, Westlaw, and PACER
PE_amt	The price erosion dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
other_dmg_flag	The patent holder was awarded other damages	PwC database, Google, Westlaw, and PACER
other_dmg_amt	The other damages dollar award granted to the patent holder for the patent claims in the trial case before appeal of damages	PwC database, Google, Westlaw, and PACER
Settlement	The case settled after a finding of validity and infringement but before damages were awarded	PwC database, Google, Westlaw, and PACER
Injunction	The patent holder was awarded an injunction, but no other damages	PwC database, Google, Westlaw, and PACER
ANDA	The case involved an ANDA filing by the potential infringer (injunction and possibly costs awarded but no other damages for patent infringement)	PwC database, Google, Westlaw, and PACER

List of Variables and Descriptions (cont.)

Variable	Description	Source
Category 2: Litigant Information		
Number_Assignees	Number of patent assignees associated with the patents-at-issue in the case	NBER patent database, Google, and Westlaw
Pat_Assignee	Name of the assignee over all patents-at-issue in the case; one variable for each assignee	NBER patent database, Google, and Westlaw
Assignee_Unassigned	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "Unassigned"	NBER patent database
Assignee_US_Non Govt	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "US, Non-government"	NBER patent database
Assignee_Non US_Non Govt	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "Non-US, Non-government"	NBER patent database
Assignee_US Indiv	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "US, Individual"	NBER patent database
Assignee_Non US Indiv	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "Non-US, Individual"	NBER patent database
Assignee_US Govt	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "US Government"	NBER patent database
Assignee_Non US Govt	At least one of the patents-at-issue in the case had an assignee in the 2002 NBER patent database coded as "Non-US Government"	NBER patent database
P_Assignee	At least one of the patent assignee(s) is the first named plaintiff in the case	calculated
D_Assignee	At least one of the patent assignee(s) is the first named defendant in the case	calculated
Patent_Manuf_Mkt_Tech	The patent holder markets or manufactures its technology covered by the patent; =1 yes; =0 no; =2 unclear	PwC database, Google, Westlaw, and PACER
P_Individual_C	The first named plaintiff is an individual	EDGAR, Manta, Hoover's Online, and Westlaw
P_Private Entity_C	The first named plaintiff is a private entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Public Entity_C	The first named plaintiff is a public entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_University_C	The first named plaintiff is a university	EDGAR, Manta, Hoover's Online, and Westlaw
P_US Government_C	The first named plaintiff is part of the U.S. government	EDGAR, Manta, Hoover's Online, and Westlaw
P_Domestic_C	The first named plaintiff is a domestic entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Foreign_C	The first named plaintiff is a foreign entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Fortune 500_2009_C	The first named plaintiff is part of the 2009 Fortune 500	Fortune 1000
P_Fortune 1000_2009_C	The first named plaintiff is part of the 2009 Fortune 1000	Fortune 1000
p_fortune_501_1K_2009_c	The first named plaintiff is listed in the Fortune 501 to 1000 in 2009	Fortune 1000
P_Subsidiary_C	The first named plaintiff is a subsidiary of a parent company	EDGAR, Manta, Hoover's Online, and Westlaw

List of Variables and Descriptions (cont.)

Variable	Description	Source
	Category 2: Litigant Information (cont.)	
P_Private Entity_Par	The first named plaintiff is a subsidiary and the parent company is a private entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Public Entity_Par	The first named plaintiff is a subsidiary and the parent company is a public entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Domestic_Par	The first named plaintiff is a subsidiary and the parent company is a domestic entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Foreign_Par	The first named plaintiff is a subsidiary and the parent company is a foreign entity	EDGAR, Manta, Hoover's Online, and Westlaw
P_Fortune 500_2009_Par	The first named plaintiff is a subsidiary and the parent company is in the 2009 Fortune 500	Fortune 1000
P_Fortune 1000_2009_Par	The first named plaintiff is a subsidiary and the parent company is in the 2009 Fortune 1000	Fortune 1000
p_fortune_501_1K_2009_par	The first named plaintiff is a subsidiary and the parent company listed in the Fortune 501 to 1000 in 2009	Fortune 1000
P_Joint Venture_Par	The first named plaintiff is a subsidiary and is owned by a joint venture	EDGAR, Manta, Hoover's Online, and Westlaw
D_Individual_C	The first named defendant is an individual	EDGAR, Manta, Hoover's Online, and Westlaw
D_Private Entity_C	The first named defendant is a private entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Public Entity_C	The first named defendant is a public entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_University_C	The first named defendant is a university	EDGAR, Manta, Hoover's Online, and Westlaw
D_US Government_C	The first named defendant is part of the U.S. government	EDGAR, Manta, Hoover's Online, and Westlaw
D_Domestic_C	The first named defendant is a domestic entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Foreign_C	The first named defendant is a foreign entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Fortune 500_2009_C	The first named defendant is part of the 2009 Fortune 500	Fortune 1000
D_Fortune 1000_2009_C	The first named defendant is part of the 2009 Fortune 1000	Fortune 1000
d_fortune_501_1K_2009_c	The first named defendant is listed in the Fortune 501 to 1000 in 2009	Fortune 1000
D_Subsidiary_C	The first named defendant is a subsidiary of a parent company	EDGAR, Manta, Hoover's Online, and Westlaw

List of Variables and Descriptions (cont.)

Variable	Description	Source
Category 2: Litigant Information (cont.)		
D_Private Entity_Par	The first named defendant is a subsidiary and the parent company is a private entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Public Entity_Par	The first named defendant is a subsidiary and the parent company is a public entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Domestic_Par	The first named defendant is a subsidiary and the parent company is a domestic entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Foreign_Par	The first named defendant is a subsidiary and the parent company is a foreign entity	EDGAR, Manta, Hoover's Online, and Westlaw
D_Fortune 500_2009_Par	The first named defendant is a subsidiary and the parent company is in the 2009 Fortune 500	Fortune 1000
D_Fortune 1000_2009_Par	The first named defendant is a subsidiary and the parent company is in the 2009 Fortune 1000	Fortune 1000
d_fortune_501_1K_2009_par	The first named defendant is a subsidiary and the parent company listed in the Fortune 501 to 1000 in 2009	Fortune 1000
D_Joint Venture_Par	The first named defendant is a subsidiary and is owned by a joint venture	EDGAR, Manta, Hoover's Online, and Westlaw
ind_SIC2	The 2-digit SIC code for the potential infringer	NBER patent database, Google, and Westlaw
ind_sic_mining	Equals 1 if *ind_sic2* is between 10 and 14 inclusive	NBER patent database, Google, and Westlaw
ind_sic_cons	Equals 1 if ind_sic2 is between 15 and 17 inclusive	NBER patent database, Google, and Westlaw
ind_sic_manuf	Equals 1 if ind_sic2 is between 20 and 39 inclusive	NBER patent database, Google, and Westlaw
ind_sic_trans	Equals 1 if ind_sic2 is between 40 and 49 inclusive	NBER patent database, Google, and Westlaw
ind_sic_whole	Equals 1 if ind_sic2 is between 50 and 51 inclusive	NBER patent database, Google, and Westlaw
ind_sic_retail	Equals 1 if ind_sic2 is between 52 and 59 inclusive	NBER patent database, Google, and Westlaw
ind_sic_finance	Equals 1 if ind_sic2 is between 60 and 67 inclusive	NBER patent database, Google, and Westlaw
ind_sic_services	Equals 1 if ind_sic2 is between 70 and 89 inclusive	NBER patent database, Google, and Westlaw
ind_sic_pubadmin	Equals 1 if ind_sic2 is between 90 and 99 inclusive	NBER patent database, Google, and Westlaw
ind_SIC3	The 3-digit SIC code for the potential infringer	NBER patent database, Google, and Westlaw
ind_SIC4	The 4-digit SIC code for the potential infringer	NBER patent database, Google, and Westlaw

List of Variables and Descriptions (cont.)

Variable	Description	Source
	Category 3: Patent(s)-at-Issue Information	
Number_Patents	Number of patents-at-issue in the case	Google, Westlaw, and PACER
Pat_Utility	One or more of the patents-at-issue are a utility patent	NBER patent database, Google, and Westlaw
Pat_Reissue	One or more of the patents-at-issue are a reissue patent	NBER patent database, Google, and Westlaw
Pat_Design	One or more of the patents-at-issue are a design patent	NBER patent database, Google, and Westlaw
Pat_Application	One or more of the patents-at-issue are an application number	NBER patent database, Google, and Westlaw
Pat_App_year	Application year of all patents-at-issue in the case calculated for minimum and maximum	NBER patent database, Google, and Westlaw
Pat_Gyear	Grant date year of all patents-at-issue in the case, calculated for minimum and maximum	NBER patent database, Google, and Westlaw
Pat_Gdate	Grant date of all patents-at-issue in the case, calculated for minimum and maximum	NBER patent database, Google, and Westlaw
pat_age_first	Age of the oldest patent-at-issue from date of complaint, calculated in days and years	calculated
pat_age_last	Age of the youngest patent-at-issue from date of complaint, calculated in days and years	calculated
pat_age_avg	Average age of all the patents-at-issue from date of complaint, calculated in days and years	calculated
Pat_Claims	Number of claims of all patents-at-issue in the case, calculated for minimum, maximum, average, and total	NBER patent database, Google, and Westlaw
Pat_Fwd_Cite_02	Number of forward citations of all patents-at-issue in the case from the NBER 2002 coding, calculated for minimum, maximum, and average number of forward citations through 2002	NBER patent database, Google, and Westlaw
Pat_Fwd_Cite_10	Number of forward citations of all patents-at-issue in the case not available in the NBER 2002 coding, calculated for minimum, maximum, and average number of forward citations through early 2010	Google and Westlaw
IPC4_Human_Nec	One or more of the patents-at-issue had an IPC code that began with "A" (Human Necessities)	NBER patent database, Google, and Westlaw
IPC4_Perf_Ops	One or more of the patents-at-issue had an IPC code that began with "B" (Performing Operations; Transporting)	NBER patent database, Google, and Westlaw
IPC4_Chem	One or more of the patents-at-issue had an IPC code that began with "C" (Chemistry; Metallurgy)	NBER patent database, Google, and Westlaw
IPC4_Textiles	One or more of the patents-at-issue had an IPC code that began with "D" (Textiles; Paper)	NBER patent database, Google, and Westlaw
IPC4_Construction	One or more of the patents-at-issue had an IPC code that began with "E" (Fixed Constructions)	NBER patent database, Google, and Westlaw
IPC4_Mech_Engineering	One or more of the patents-at-issue had an IPC code that began with "F" (Mechanical Engineering; Lighting; Heating; Weapons; Blasting)	NBER patent database, Google, and Westlaw
IPC4_Physics	One or more of the patents-at-issue had an IPC code that began with "G" (Physics)	NBER patent database, Google, and Westlaw
IPC4_Electricity	One or more of the patents had an IPC code that began with "H" (Electricity)	NBER patent database, Google, and Westlaw
PTO_Main_Class	PTO Main Class Code for patent-in-suit; each individual patent in the case has its own variable	NBER patent database, Google, and Westlaw