REDEFINING "VALUABLE PATENTS": ANALYSIS OF THE ENFORCEMENT VALUE OF U.S. PATENTS

Jonathan H. Ashtor*

CITE AS: 18 STAN. TECH. L. REV. 497 (2015)

ABSTRACT

This study analyzes the factors that make patents valuable in enforcement awards. Leading scholarship predominantly relies on proxies for value (e.g., whether a patent has been asserted or maintained), to designate which patents are "valuable" and which are not. Here, we study value directly and precisely, identifying the specific characteristics that are associated with higher or lower monetary enforcement values. In so doing, we identify previously unobserved characteristics of "valuable patents" and their values in litigation.

Specifically, we mine a vast array of data relating to each patent that has been held valid and infringed and for which damages have been awarded in U.S. District Court cases from 2006 to 2011. The dataset comprises nearly 400 patents from over 200 cases awarding infringement damages during this six-year timeframe. We use the damages awarded for infringement as an exact quantitative measure of value, which we analyze with reference to over 70 unique data points for each patent, including variables regarding prosecution history, inventors, specification and claim structure, family tree, forward citations and recorded transfers and liens.

Based on our analysis, we categorize "valuable patents" from the perspective of enforcement awards as follows: (1) commercialized patents, (2) upstream patents, and (3) forward-cited patents, with certain caveats. We further investigate each of these categories to provide new insights into patent enforcement value and articulate relevant distinctions between enforcement and other types of patent value, such as licensing or transfer value.

The research and writing of this work was supported by a Leonardo da Vinci Fellowship awarded by George Mason University School of Law.[1]

---

* Associate, Skadden, Arps, Slate, Meagher & Flom LLP. The views expressed herein are solely those of the author and do not reflect the views of others, including PricewaterhouseCoopers LLP, Skadden, Arps, Slate, Meagher & Flom LLP (or its attorneys or clients), or any of their respective affiliates.

1. The author is grateful for the foundational support and guidance of Adam Mossoff, Mark Schultz, and the Center for the Protection of Intellectual Property. The author is also grateful to Larry Ranallo, Christopher Barry, Ronan Arad, and PricewaterhouseCoopers LLP

TABLE OF CONTENTS

for the proprietary database on which this study was built. Thanks are also due to the organizers of the 14th Annual Intellectual Property Scholars Conference at Boalt Hall in Berkeley, Calif. on Aug. 7-8, 2014, where preliminary results of this study were presented. Additionally, this work has benefited from the helpful comments and contributions of many, including Michael Mazzeo, David Schwartz and Samantha Zyontz. Finally, the author thanks the tremendous research work and feedback provided by Rebecca Cress, Scott Flanz, Jordan Joachim, Megan Manfred, Joanie Piasta, Josh Podolnick, Tina Praprotnik, Hoda Rifai-Bashjawish, Aliya Sanders, and Katie Webb, as well as the entire Patent Analytics Team at Skadden.

INTRODUCTION

"Valuable patents" as the term is defined in leading scholarship are generally understood to be those patents that have been asserted in litigation or maintained to full term. However, these proxies for value are over-inclusive, and they further fail to specify *how much* value a particular patent has. Moreover, proxy analysis often lacks the precision to determine what factors make one patent more or less valuable than another. In light of the tremendous importance of patents as an asset class and the modern economy of patent enforcement, licensing and other monetization activities, it is imperative that we develop an accurate understanding of patent value and its principal characteristics.

This study analyzes patent enforcement value directly, focusing on the amount of damages awarded in court for infringement of individual U.S. patents. Our dataset comprises U.S. District Court patent infringement awards from 2006 to 2011, including nearly 400 patents that were held valid and infringed in over 200 cases decided during this period. Moreover, we compile one of the most comprehensive arrays of intrinsic and acquired patent attributes to date, including over 70 variables coding the features of each patent, characteristics of the litigants, and case factors. By applying regression analysis and other statistical techniques to this dataset, we parse out three distinct facets of patent enforcement value and further identify distinguishing characteristics that are associated with each.

Specifically, we find that the patents most valuable in enforcement proceedings are (1) commercialized patents, (2) upstream patents, and (3) with certain caveats, forward-cited patents. First, our analysis reveals that one of the most important measures of patent enforcement value is whether the claimed invention is being commercialized by its owner, and in turn we find evidence that the amount of damages likely to be awarded is strongly dependent on the market economics in which such commercialization occurs. Second, we observe a significant value premium for upstream patents and differentiation between upstream and downstream patents, whereby "upstream" patents that give rise to follow-on inventions and improvements have significantly higher value than other patents on average and, in particular, relative to patents for those "downstream" derivatives. Furthermore, we find that forward citations strongly correlate with patent enforcement value; however this signal is somewhat less clear, and the presence of strong relationships between forward citations and many other patent characteristics, as well as certain deviations at the high value range, suggest a more complicated relationship between forward citations and value.[2] Additionally, we find that proprietary patents that have been held and enforced by their original applicants are significantly more valuable in court awards than patents acquired from third parties prior to enforcement, which further reinforces the significance of the commercialization and development functions we observe.

Importantly, we note at the outset that our findings regarding patent enforcement value may not be generalizable to other types of patent value, particularly given the selection of our dataset. We are exclusively studying fully litigated cases that have not settled. Moreover, we do not address licensing or other monetization activities that are undertaken without litigation. There are many reasons to think that different factors may drive patent value in different settings depending on how the patent is being used, and therefore it is important to be mindful of the particular context of any study of patent value.

Nonetheless, we attempt here to provide a detailed picture of patent enforcement value in court awards and explore its various dimensions. In the analysis below, we investigate patents enforced by non-practicing entities, patents involved in large-entity litigations, patents in different industry sectors and patents that have been traded, in each case controlling for a wide range of patent characteristics. We also pose for future study certain key questions about forward citations, which despite being the most prominent metric for estimating patent value have certain key limitations and complicating factors.

This study is organized as follows. In Part II below, we outline relevant theoretical background and prior scholarship. Next, Part III describes our dataset and empirical methodology, including detailed descriptions of the variables we code and summary statistics and trends pertaining to our data of patent awards. Part IV details our correlation analysis, large-scale regression modeling and specific investigation of key parameters, and provides our results. Finally, our interpretations and conclusions follow in Part V.

I. THEORETICAL BACKGROUND

Most prevailing theories of patent value define "valuable patents" as those

---

2. *See infra* Part IV.C. We also compare our results with other work that has found complexities and a non-monotonic citation to value relationship. *See* Abrams et al., *infra* note 33.

likely to be asserted in infringement litigation.[3] However, assertion is an imprecise proxy for value—for instance, most completed patent litigations result in non-infringement or invalidity findings (*i.e.*, no value), and most of those that find infringement result in relatively low damages awards (low value). Also, the vast majority of patent lawsuits are settled before trial, and these may similarly result in very low value or, in cases where the patent holder drops the suit, no value. Yet, assertion-based proxies will label any asserted patent as "valuable" irrespective of the actual amount of value, if any, eventually awarded. Other widely-used proxies focus on whether a patent has been maintained or abandoned, but these similarly do not reflect the fact that most patents maintained as active are never licensed or asserted or even practiced, nor do these proxies reflect the vast disparity in values of those few maintained patents that have any net worth.

Moreover, different sources of patent value are likely to vary widely depending on how the patent is used, whereas the general term "patent value" overlooks such distinctions. Value from licensing revenues versus firm market capitalization, and intangible asset worth versus enforcement value, are not necessarily directly comparable to each other, and different factors may have different influences on each type. For example, a patent that has been widely licensed may generate low damages in an individual litigation against an accused infringer,[4] and a patent held for defensive purposes may have high value to its owner precisely because it is not licensed to third parties. Even specifically focusing on value in litigations, some patents may be more valuable in voluntary settlements relative to fully-litigated enforcement awards (*e.g.*, NPE patents versus practiced patents, where the latter may be more likely to garner higher awards of lost profits).[5]

Here, we focus on patent enforcement value via the amounts awarded for infringement of an individual patent in fully litigated cases, which allows us to precisely examine the features of this type of patent value and distinguish it from other types. The sections below describe the leading prior scholarship, focusing on the predominant proxy-based studies in Part A, and addressing in Part B on the few direct studies of value that have been conducted. This section concludes by describing patent enforcement value as a specific type of patent value and identifying prior work relevant to the study thereof.

---

3. *See, e.g.*, John R. Allison et al., *Valuable Patents*, 92 GEO. L.J. 435, 449 (2004).

4. In particular, prior license royalties are typically taken into account when computing reasonable royalties. *See generally ResQNet.com, Inc. v. Lansa, Inc.*, 594 F.3d 860 (Fed. Cir. 2010).

5. *See* Michael J. Mazzeo et al., *Do NPEs Matter? Non-Practicing Entities and Patent Litigation Outcomes*, 9 J. COMPETITION L. & ECON. 879 (2013) [hereinafter Mazzeo et al., *Do NPEs Matter?*] (analyzing the characteristics of NPE patents and finding that NPEs are likely to receive lower award amounts relative to other patent holders).

A. *Proxies for Value*

The leading studies of patent value use proxies for economic value and seek to empirically determine the factors that influence the proxy in question. A prime example is *Valuable Patents* by Allison, Lemley et al.,[6] which focuses on whether a particular patent will be asserted in court. The authors argue that the connection between assertion and core asset value is strong, but they also acknowledge the limitations of their proxy: "While not every valuable patent is necessarily litigated, we believe that the relationship is strong enough to justify the conclusion that litigated patents are a good proxy for valuable patents. . . . We acknowledge that the litigation/value connection is an important and controversial assumption."[7]

The authors conclude that, relative to non-litigated patents, "valuable patents" according to their definition are more likely to: (1) be younger, (2) be owned by domestic rather than foreign companies, (3) issue to individuals or small companies, (4) have more forward and backward citations, (5) spend longer in prosecution, (6) contain more claims and (7) predominantly come from the mechanical, computer, medical device, and other select industries.[8] Notably, the authors also observe that "valuable patents" tend to have more continuations and other child applications than non-litigated patents[9] or otherwise come from large patent families (*i.e.*, clusters of domestic patents deriving from the same ultimate parent).[10] However, the authors do not expressly distinguish between upstream and downstream patents or find significant differences in value between them.[11]

The *Valuable Patents* article spawned a number of follow-on and responsive studies and commentaries by both the original authors and other academics. For example, Allison, Lemley, and Walker also studied "extreme value" patents, which they define as those patents asserted in multiple litigations, and further focusing on the role of non-practicing entities in launching multi-assertion campaigns.[12] Also, Miller analyzed the connection between these most-litigated patents and validity.[13] And Chien has further

6. Allison et al., *supra* note 3.

7. *Id.* at 443.

8. *Id.* at 438.

9. *Id.* at 456-8.

10. *Id.* at 457 n.93.

11. Rather, both patents that give rise to more children and patents with larger overall families are found to be "valuable" based on likelihood of assertion. *Id.* at 457-58 ("Litigated patents were part of a family of 1.85 patents on average, while non-litigated patents had a family size of only 1.22.").

12. *See generally* John R. Allison et al., *Extreme Value or Trolls on Top? The Characteristics of the Most Litigated Patents*, 158 U. PA. L. REV. 1 (studying litigation rates of patents in specific industries).

13. *See generally* Shawn P. Miller, *What's the Connection Between Repeat Litigation and Patent Quality? A (Partial) Defense of the Most Litigated Patents*, 16 STAN. TECH. L. REV. 313 (2013).

extended the methodology of predicting patent litigation, including by introducing new "intrinsic" and "acquired" patent attributes, such as recorded transfers and liens.[14]

Valuable Patents and its progeny draw upon a formative set of articles by Lanjouw and Schankerman, which study the predictors of patent infringement suits in a broader economic context including market and industry factors, litigant characteristics, sector-specific patent densities and technology class.[15] These studies identify certain characteristics of parties and their patent assets that increase the likelihood of a suit being filed in a particular market/industry context. They find in part that the probability of patent litigation increases with respect to patents that are central to follow-on innovations of a company, particularly between companies that are close rivals or where the patent holder needs to maintain a reputation for aggressive enforcement.[16] By contrast, companies in concentrated industries or with particularly large patent portfolios relative to others are less likely to engage in litigation as they often have other means of avoiding disputes or cross-licensing.[17] These studies further identify certain specific patent characteristics that increase the probability of assertion, most notably patents having a higher number of claims and more forward citations per claim.[18] Notably, the authors do not study these attributes in isolation, instead analyzing them relative to the distributions of similarly-situated patents based on industry classification and other groupings.[19]

Finally, another common proxy for value focuses on patent "mortality" rates, namely the likelihood that an entity will abandon its patent rather than pay scheduled maintenance fees to the USPTO at statutory intervals. Like assertion, maintenance is a reasonable proxy for value in many respects—if a patent holder pays to maintain a patent then it is likely to be more valuable to that entity than an abandoned patent. However, payment of maintenance fees is

---

14. Colleen V. Chien, *Predicting Patent Litigation*, 90 TEX. L. REV. 283, 298-99 (2011).

15. *See, e.g.*, Jean O. Lanjouw and Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129 (2001); Jean O. Lanjouw and Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, 47 J.L. & ECON. 45 (2004); Jean O. Lanjouw and Mark Schankerman, *Research Productivity and Patent Quality: Measuring Innovation with Multiple Indicators*, 114 ECON. J. 223, 441-65 (2004).

16. Lanjouw and Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, *supra* note 15, at 129-30.

17. Lanjouw and Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, *supra* note 15, at 48.

18. Lanjouw and Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, *supra* note 15, at 131.

19. *Id*. Implicitly this approach acknowledges that forward citations are influenced by contextual factors—*e.g.*, patents in more densely-patented fields are more likely to have a higher number of forward citations arising during ordinary prosecution of subsequent patents by other applicants. Although this supports the connection between forward citations and assertion (since patent density is a strong predictor of assertion), the connection to economic value is less direct.

a very broad brush for tracing the outlines of patent value, given that renewal fees are quite low and companies have a variety of reasons for maintaining their portfolios independent of specific value assessments (including common contractual requirements to maintain assets that are subject to blanket financing or license agreements).

One study in this tradition, authored by Barney in 2002,[20] identified several statistical markers for patent "value" as represented by the survival-mortality relationship. Specifically, that study found that maintained patents are more likely to come from certain IPC classes, including that "patents relating to genetic engineering and computers appear to be statistically more valuable."[21] It also analyzed structural features, finding that patents having more independent claims, shorter claims, and longer specifications are more valuable.[22] Interestingly, it also found that patents claiming priority to early applications are likely to be *more* valuable, explaining that: "Intuitively, more priority claims probably means a patent is entitled to an earlier filing date, which can be beneficial in fending off art-based validity attacks. It could also indicate a greater level of overall interest and investment by the patentee."[23] The study does not appear to examine continuity data or child applications, or make the distinction between upstream and downstream patents. Finally, it finds a strong correlation between forward citations and patent value, echoing prior studies that reached similar conclusions.[24]

B. *Direct Studies of Value*

Certain studies have taken a different approach than the value-proxy corpus and analyzed economic value directly. These studies often use market capitalization of the patent holders as their measure of value, although some more recent studies have focused on royalties from certain types of licensing relationships,[25] although none focus on enforcement value from infringement awards.

---

20. Jonathan A. Barney, *A Study of Patent Mortality Rates: Using Statistical Survival Analysis to Rate and Value Patent Assets*, 30 AIPLA Q.J. 317, 329 (2002). Another study examining patent renewal rates was authored by Bessen in 2006. James Bessen, *The Value of U.S. Patents by Owner and Patent Characteristics* (Boston Univ. Sch. of Law, Working Paper No. 06-46, 2006), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=949778. These studies also draw on foundational scholarship by Mark Schankerman and Ariel Pakes. *See*, *e.g.*, Ariel Pakes & Mark Schankerman, *The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Return to Research Resources*, *in* R&D, PATENTS, AND PRODUCTIVITY 73-88 (Zvi Griliches ed., 1984); Mark Schankerman & Ariel Pakes, *Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period*, *in* 96-384 THE ECONOMIC JOURNAL, 1052-1076 (1986).

21. Barney, *supra* note 20, at 331.

22. *Id*. at 332-33.

23. *Id*. at 333. *Cf. infra* Part IV.C (finding that priority claims are significantly *negatively* correlated with award value).

24. *Id*. at 333 n.43.

25. *See* Abrams et al., *infra* note 33.

One of the leading studies that analyzes market capitalization value was published in 2005 by Hall, Jaffe, and Trajtenberg, titled *Market Value and Patent Citations*.[26] This work provides conceptual and practical foundations for studying forward citations generally, and it specifically investigates the extent to which citations of a firm's patents collectively are associated with that firm's market valuation, as measured by Tobin's Q component ratios of R&D to assets, patents to R&D and forward citations to patents. Their work further developed statistical techniques to correct forward citation counts for age truncation effects (*e.g.*, the fact that forward citations accrue over time and are therefore dependent on the measurement date and likely to be highly truncated for young patents).[27] This study also accounts for the distributional significance of different technology classes—as with Lanjouw and Schankerman's methodology described above, they adjust for expected lifetime citation counts relative to the patenting density of the particular industry sector.[28] The authors further study specific details of forward citations that are associated with higher market value, finding that high citation concentrations correspond to significant market premiums and further observing significant differences between self-citations (by subsequent applications of the same patent holder) versus citations by third party applications.[29]

Notably, market capitalization based on portfolio value is still a proxy for individual patent value (although arguably a more precise one than assertion or mortality). Furthermore, one might expect forward citations to be more strongly correlated with the market value of a firm's knowledge stock than with patent-specific enforcement award value, since market value and forward citations operate to a certain extent by similar mechanisms. That is, both function as independent estimations of the worth or importance of an underlying asset, rather than solely as private owner-informed valuations of the asset itself.[30]

Importantly, our results herein are consistent with certain key findings of Hall et al., which lends credibility to our analysis. Specifically, they find that self-citations from "down-the-line patents owned by the same firm" generally indicate higher value than third party citations.[31] This is consistent with our

26. Bronwyn H. Hall, Adam Jaffe & Manuel Trajtenberg, *Market Value and Patent Citations*, 36 RAND J. ECON. 16 (2005).

27. *Id*. at 21-22.

28. *Id*.

29. *Id*. at 31-33. *See also* Bronwyn H. Hall, Adam Jaffe & Manuel Trajtenberg, *The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools* (Nat'l Bureau of Econ. Research, Working Paper No. 8498, 2001), http://www.nber.org/papers/w8498.pdf.

30. One might also expect certain network effects in forward citations, whereby patents that are highly cited may be more likely to register or be taken seriously in prior art searches and therefore be cited again by subsequent applications. Moreover, there may be feedback effects for market value in that forward citations are publicly reported and therefore may influence market value directly to the extent they are used in analysts' assessments and investment algorithms.

31. Hall et al., *supra* note 26, at 31-33.

finding that upstream patents are significantly more valuable than downstream patents—although defined in terms of continuation applications rather than "down-the-line patents" which generate self-citations, both upstream patents and self-cited patents are defined by their impact on follow-on inventions by their owners.[32]

Transitioning from market value to licensing value, a recent study of the connection between licensing value and forward citations was conducted by Abrams, Akcigit, and Popadak, titled *Patent Value and Citations: Creative Destruction or Strategic Disruption?*.[33] The authors utilize a proprietary dataset of licensing revenues derived from patents owned by a large non-practicing entity and analyze the relationship between forward citations and lifetime licensing value. Although their dataset is somewhat selective, their results are striking. In particular, they model the citation-licensing value relationship as an inverted U-curve, finding that citations tend to increase with value for "productive" innovation efforts aimed at developing and patenting new technologies, whereas forward citations decrease with value for "strategic" patents aimed at protecting established market share and preventing entry by subsequent firms.[34]

Their findings also provide support for the upstream-downstream distinction observed herein: in particular, they find evidence that continuation and divisional patents (*i.e.*, downstream patents) are more likely to be associated with "strategic" patenting which individually may have lower value, whereas the earlier upstream patents that are protected by such efforts may have higher value and sustain their market prominence longer as a result.[35] Yet, they also predict that strategic upstream patents are likely to have a lower citations-to-value relationship, and they observe strong negative quadratic terms in their citation-to-value regressions that give rise to the inverted-U curve they report. We conduct specific analyses below to investigate the extent to which their findings from NPE licensing value are consistent with enforcement value in court awards.

Importantly, there are many conceptual reasons why licensing value may not be directly comparable to enforcement value. Rather, voluntary transactions differ from court enforcements in several respects. Even to the extent licenses are negotiated in the "shadow of litigation," voluntary transactions often involve sources of value apart from the patents themselves (including associated services, information exchange, etc.).[36] Also, other considerations

32. *See infra* Part IV.C.

33. Abrams et al., *Patent Value and Citations: Creative Destruction or Strategic Disruption?* 1 (Nat'l Bureau of Econ. Research, Working Paper No. 19647, 2013), http://www.nber.org/papers/w19647.

34. *Id.* at 2.

35. *Id.* at 7.

36. Indeed, courts recognize the difficulty in using licensing royalties as a benchmark for litigation outcomes, and the selection of comparable licenses and extent to which their economics impact reasonable royalty awards are often heavily litigated.

often influence negotiations, such as relative bargaining power, the reputations of the parties, and the desire to avoid the high costs of patent litigation. More fundamentally, voluntary transactions including settlements reflect the "meeting-of-the-minds" between parties as to whether and on what terms on which to license a given patent, whereas enforcement awards represent the special case where parties do not agree to any license.

In this light, enforcement awards provide a unique measure of patent asset value relative to other measures. By minimizing transactional incentives and stock market influences that may affect market-based measures, and by avoiding the imprecision of proxies, court awards offer a specific monetary valuation of the patent rights held to be infringed. We refer to this measure as *patent enforcement value*,[37] which arises from litigating a claim to judgment and obtaining a non-negotiated remedy.[38]

Prior work by the author and two coauthors provides a framework for analyzing patent enforcement value.[39] Those studies conducted large-scale empirical analyses of patent infringement awards, focusing on the factors that are associated with the aggregate value awarded in each case. The results identified key predictors of patent case value, including certain characteristics of the parties (such as enterprise size and whether the plaintiff is an NPE), type of award (such as lost profits versus reasonable royalties) and relevant patent attributes. However, those studies do not address patent-specific value (as opposed to case value) or focus in depth on the individual patent attributes that are associated with such value. The present study builds from that framework to analyze patent-specific enforcement value and the attributes associated therewith, using the methodology described below.

II. DATASET AND METHODOLOGY

This Part describes our empirical methodology. Specifically, Part A describes the dataset and data collection procedure used to code it. Part B details the preprocessing methodology for refining the dataset and preparing the variables for statistical analysis. Finally, Part C provides summary statistics and distribution analyses of the final data.

---

37. More specifically, patent enforcement value as measured by lost profits or reasonable royalties represents the amount of compensatory damages for infringement proven to and awarded by a court within the market and legal framework existing at the time.

38. We do not claim that enforcement value is independent of market factors or transaction value. To the contrary, we argue that enforcement value provides a clear quantitative measure from which to study the influence of various patent characteristics, and from which to derive models of value for further study.

39. Michael J. Mazzeo et al., *Explaining the "Unpredictable": An Empirical Analysis of U.S. Patent Infringement Awards*, 35 INT'L REV. L. & ECON.58 (2013) [hereinafter Mazzeo et al., *Explaining the "Unpredictable"*]; *see also* Jonathan H. Ashtor et al., *Patents at Issue: The Data Behind the Patent Troll Debate*, 21 GEO. MASON L. REV. 957 (2014); Mazzeo et al., *Do NPEs Matter?*, *supra* note 5.

A. *Dataset*

Our dataset comprises U.S. District Court cases decided from 2006 through 2011 in which a monetary award was granted for patent infringement. We start from a database of all patent cases reported in Westlaw from 1995-2011, which is maintained and was licensed to us by PricewaterhouseCoopers ("PwC").[40] Specifically, our final dataset comprises a total of 221 cases and 385 patents, with over seventy unique variables for each patent. The following paragraphs provide additional detail regarding the cases, patents, and variables studied herein.

1. *Cases and Patents*

The complete PwC dataset from 1995 through 2011 contains 1,751 patent cases in Westlaw in which a decision was made on patent validity and infringement, either at summary judgment or following trial. Of those 1,751 cases, 421 included holdings of validity and infringement for at least one patent and either had available award amounts or were cases related to Abbreviated New Drug Application (ANDA) litigation. There were 45 ANDA cases with $0 awards (since ANDA cases do not result in damages) and 376 cases with reported awards greater than $0.

In the target period of 2006-2011, the PwC dataset contains a total of 335 patent cases in which at least one patent was held valid and infringed. Of these cases, forty were ANDA cases and seventy-four did not have a reported award (likely due to post-trial settlement before damages were assessed). The awards in the remaining 221 cases comprise the dataset used herein.[41]

After selecting the relevant years from the PwC dataset, our research team retrieved from Westlaw the complaints, dockets, opinions and verdict forms of each case. Using these materials, we verified the award amount and identified the patent(s) on which the award was based. To ensure reproducibility of our results and avoid selection bias, we did not add or exclude any cases from the dataset during this phase of coding; unlike proprietary datasets based on unpublished or limited data, our dataset should reflect the majority of reported decisions during the relevant time period.[42]

---

40. *See 2014 Patent Litigation Study*, PRICEWATERHOUSECOOPERS (July 2014), http://www.pwc.com/en_US/us/forensic-services/publications/assets/2014-patent-litigation-study.pdf [hereinafter 2014 PwC Study]. PwC uses this data in its annual reports and statistics on patent litigation, which are widely cited and used by academics, practitioners and government policy-makers.

41. The breakdown of cases per year are as follows:

2006	2007	2008	2009	2010	2011
33	39	45	42	36	26

42. We note that other sources report higher case counts for this time period, such as PatStats.org. Certain discrepancies may be due to different categorizations of infringement awards (*e.g.*, we do not include contractual disputes), awards for particular types of claims

2. *Intrinsic Attributes*

To analyze the impact of structural and other inherent patent characteristics on award value, we coded approximately forty-five "intrinsic attributes" for each patent in the dataset.[43] Intrinsic attributes are useful in the analysis because they exist at the time the patent is granted and, with certain exceptions,[44] remain unchanged through the life of the patent. Accordingly, they generally are unaffected by *ex post* events such as patent transfers, collateralizations, enforcement, and licensing and are not influenced by product-specific and market-driven effects, such as the ultimate commercialization value of the patented technology.

The intrinsic attributes studied herein can be loosely grouped into the following four categories: priority attributes, structural features, prosecution history and backward citations, and inventor characteristics. Each category is further described below, and the complete list of intrinsic patent attributes included in our dataset is provided in Appendix A.

a. *Priority Attributes:*

The variables in this category relate to priority claims made by the applications from which the asserted patents issue, including whether the application was published prior to issuance, whether the application claimed priority to an earlier provisional or non-provisional application, whether the application was a direct child of another application (*e.g.*, a continuation, continuation-in-part or divisional), and the number of parent applications identified on the face of the patent. We also coded various dates associated with these factors, such as the original priority date and the filing date, from which we calculated the duration that elapsed between them. These variables were coded from USPTO records and the patents themselves together with information from Public PAIR and Google Patents.

(*e.g.*, we exclude all ANDA cases) and damages awarded (*e.g.*, we do not count fees or other monetary awards that are not directed to compensation for infringement). Nonetheless, there are likely some reporting gaps, although we expect that these are likely to be small and generally unbiased. We are grateful to David Schwartz for pointing out the discrepancy with the PatStats counts.

43. The "intrinsic" versus "acquired" attributes terminology is commonly used in prior scholarship. *See*, *e.g.*, Chien, *supra* note 14, at 287.

44. Notably, intrinsic characteristics are not entirely uninfluenced by *ex post* events. For example, structural features, such as the number of dependent and independent claims, may be changed through reexamination proceedings during the life of the patent, which are more likely to be initiated when a patent is asserted. We coded the intrinsic attributes existing as of the time of assertion, accounting for changes due to reexaminations, and we included a flag indicating whether the patent was reexamined.

b. *Structural Features*

These variables relate to the type and structure of the patents asserted, such as the application type (utility, design, or plant), first industry classification (IPC category),[45] number of dependent and independent claims, respective lengths of the written description and claims (measured as number of patent columns of each), number of figures, and boolean flags for whether the patent contains formulas or data tables. These variables were coded from the patent files.[46]

c. *Prosecution History and Backward Citations*

We gathered extensive information relating to the prosecution history of the patents prior to issuance and citations to prior art made by the applicant and examiner. These variables include the duration of prosecution, number of Office Action rejections, and boolean flags for terminal disclaimers, term adjustments, and certificates of correction. Also, we counted the numbers of backward citations made by the applicant and the examiner, respectively, as well as the type of prior art cited (patents versus non-patent references). These variables were largely coded from Public PAIR records and the patent documents.

d. *Inventor Characteristics*

We also coded several variables to investigate the relationship between the inventors of a patent and its expected enforcement value. Specifically, we counted the number of total inventors and specific numbers of domestic and foreign inventors named on each patent (based on their reported country of residence). Additionally, using Public PAIR we also coded whether the applicant filed with small entity or undiscounted status.[47]

3. *Acquired Attributes:*

As discussed above, prior scholarship has recognized the importance of acquired attributes,[48] and we included several acquired attributes in our dataset to investigate the relationship with enforcement value. Specifically, our dataset

---

45. *See infra* Appendix D.

46. Other studies have included structural features in analyzing proxies for value. *See*, *e.g.*, Barney, *supra* note 20, at 332-33. Also, recent work has used sophisticated computational analysis of structural features to analyze validity. *See*, *e.g.*, Yan Liu et al., *Latent Graphical Models for Quantifying and Predicting Patent Quality*, 17 ACM SIGKDD INT'L CONF. ON KNOWLEDGE DISCOVERY AND DATA MINING 1145 (2011).

47. Micro-entity status was not relevant for the patents in our dataset given the date range.

48. *See generally* Hall et al., *supra* note 26 (studying forward citations).

includes forward citations, recorded transfers and liens, and the number of child applications claiming priority to the patent at issue, as further described below. The complete list of acquired attributes and corresponding data sources is provided in Appendix B.

a. *Forward Citations*

Using the "Referenced By" feature in the USPTO database,[49] we counted the number of citations received from other patents that were applied for as of the date of the complaint in which the patent at issue was asserted. To mitigate age-truncation bias, we then computed an age-adjusted number of forward citations for each patent by dividing by the fraction of the patent's twenty-year life elapsed as of the complaint date. This provides a first-order correction to the truncation bias, which approximates a uniform citation distribution over time (and therefore likely overstates the number of lifetime forward citations, particularly for young patents). However, employing a distribution-specific approach, such as the widely used adjustment developed by Hall, Jaffe and Trajtenberg in their NBER work,[50] would also entail significant noise given the relatively young age of the patents in our dataset relative to the population at large. In particular, we queried the 2012 updated NBER dataset, which includes all patents issued through 2006,[51] and found that 358 of the 385 total patents in our sample were included therein. Yet, many of these 358 patents had issued within a few years of 2006, meaning that the future patents citing to them are not reflected in the NBER dataset, leading to serious truncation. Accordingly, NBER's age-adjustment factors and generality measures are not necessarily reliable for the patents we study here. After employing our first-order adjustment, significance testing reveals that the strong negative correlation between forward citations actually received and age is eliminated.

Also, we note that most of the patents in our dataset lie towards the end of the 3-7 year age range (the median patent age of our sample is 7.3 years and the average is 8.3 years) when the bulk of lifetime citations are expected to be received,[52] and therefore we expect that even the unadjusted citation data is likely to capture the major proportion of lifetime citations expected to be received. Additionally, to further mitigate age truncation effects, we re-coded the forward citation counts by relaxing the limitation to only citing patents that had been applied for as of the relevant complaint date and instead including all citing patents issued through December 2014. This expands the citations

---

49. *USPTO Patent Full-Text and Image Database*, U.S. PATENT & TRADEMARK OFFICE (last visited May 7, 2015), http://patft.uspto.gov/netahtml/PTO/search-bool.html.

50. Hall et al., *supra* note 26, at 21-22 (providing a more complex adjustment technique).

51. NATIONAL BUREAU OF ECONOMIC RESEARCH, PATENT DATA PROJECT, https://sites.google.com/site/patentdataproject/Home (last visited on Oct. 24, 2015).

52. *See* Alan C. Marco, *The Dynamics of Patent Citations* 5-6 (Vassar College Economics, Working Paper No. 84, 2006).

window to encompass the major portion of the lifespan of most patents in our sample, if not the entire term in many cases. Our results are largely invariant to the different measures we employ.[53]

b. *Recorded Transfers*

We also code recorded patent assignment and transfer data,[54] in order to examine the relationship between patent transfers and value. Theoretically, one might expect a positive correlation between transfers and enforcement value, in that a patent that has been sold is one that was perceived as valuable by the transacting parties and for which the buyer's price quantifies the amount of value exchanged.[55] However, one might also expect a negative relationship on the theory that a patent holder is less likely to sell a patent with high enforcement value, particularly if it covers important products or technologies. Additionally, non-practicing entities complicate the relationship, in that they typically acquire their patents from other entities prior to enforcement and typically will record this transfer in order to demonstrate standing to enforce the patent (thereby increasing the number of recorded assignments of their patents). NPEs have been found to receive lower compensatory damages awards on average,[56] which could tend towards a negative correlation with the number of transfers. Our transfer data was coded from the USPTO Assignments database, filtering for records that pertain to third-party transfers and excluding name changes or other transactions involving the same patent holder.

c. *Recorded Liens*

We also study data on recorded liens, which have received less attention in

53. We also intend to further examine correlations using more specific citation measures, such as self-citations versus third party citations and within-field versus external citations, in follow-on studies.

54. Previous studies have also used these acquired characteristics. *See*, *e.g*., Chien, *supra* note 14, at 287. Although USPTO records only reflect transfers that have been recorded, whereas some transfers may remain private, recordation is inexpensive and the patent recording statutes protect those who promptly record against subsequent title claims. *See* 35 U.S.C. § 261 ("An assignment, grant or conveyance shall be void as against any subsequent purchaser or mortgagee for a valuable consideration, without notice, unless it is recorded in the Patent and Trademark Office within three months from its date or prior to the date of such subsequent purchase or mortgage.").

55. Notably, Serrano has empirically studied the relationship between patent assignments and value (as proxied by forward citations) and found that citations significantly correlate with assignments. *See* Carlos J. Serrano, *The Dynamics of the Transfer and Renewal of Patents*, 41 RAND J. ECON. 686 (2010). We observe significant correlations between forward citations and assignments, but we find that assignments are significantly negatively related to the actual enforcement value awarded in court. This further highlights the distinctions between types of patent value.

56. *See* Mazzeo et al., *Do NPEs Matter?*, *supra* note 5, at 899-900.

prior scholarship. Liens or security interests may be recorded for a variety of reasons, most frequently where the patent is being used as collateral for a bank loan or other financing. Theoretically one might expect a positive relationship between recorded liens and enforcement value because most financial lenders require borrowers to seek prior approval before enforcing the patents that are granted as collateral; accordingly, the patent holder may expect higher enforcement value for encumbered patents because they have taken the extra step of obtaining lender approval prior to enforcement.[57] We coded the number of security interests recorded in the USPTO Assignments database, filtering for recordations of relevant security interests and excluding releases of previously granted liens. Next we converted this data to a boolean flag of whether any liens were recorded, given that the counts were highly skewed (*i.e.*, most patents in the dataset had zero recorded liens, and only approximately 13% had a one or more liens recorded).[58]

d. *Child Applications*

Finally, we also investigated each patent's progeny of subsequent U.S. continuations and divisionals based on continuity data from Public PAIR.[59] One might expect child continuity data to be an important indicator of enforcement value, whereby patents with more children may be more likely to be early-stage inventions and thereby possibly have greater value. Also, subsequent child applications may reflect ongoing investment and R&D by the patent holder in that particular technology, which could both generate more value and also reflect that the underlying technological resource is inherently more valuable. Furthermore, prior research has linked continuation practice with strategic efforts by the patent holder to shield the invention from subsequent disruptive market entry, which may enable the technology owner to maintain market power longer and thereby accrue more enforcement value.[60]

Given the high skew of this variable (most patents had zero children, and several with children had multiple children),[61] we used a boolean flag

57. One might also expect a negative relationship if the patent holder is less likely to pledge as collateral patents that it perceives as valuable, in order to avoid encumbering them. However, given that many financing agreements require blanket liens on all patents owned by the patent holder, the patent holder is unlikely to be able to select which patents to pledge as collateral.

58. Notably, this data likely includes some reporting bias in that liens are not required to be recorded prior to enforcement; however, given that financial lenders have incentives to perfect their interests and recording against U.S. patents is a relatively cheap and routine practice for many lenders, we expect this effect to be minimal.

59. This analysis could be extended to include foreign child patents based on WIPO data, and one could expect that this would further reinforce the effects observed by domestic data alone.

60. Abrams et al., *supra* note 33, at 7.

61. Specifically, 121 patents had zero children, 78 had one child, 47 had two children, 70 had three or four children, and 69 had five or more children.

indicating whether or not subsequent U.S. child applications had been filed from each patent.

4. *Case and Party Characteristics:*

In addition to intrinsic and acquired patent attributes, our dataset includes approximately twenty variables relating to each case and the litigants. Case variables include the Circuit in which the U.S. District Court was located, flags indicating the type of damages awarded (*e.g.*, lost profits or reasonable royalties, and whether enhanced damages were awarded), whether a judge or jury decided the case, and the length of the litigation. Litigant variables include the size, industry SIC code, Fortune 500 status, and foreign versus domestic domicile of the plaintiff and defendant.[62] We principally used the complaint and final court opinion or order to code the case-related variables, and we used Mergent records and the Forbes Fortune 500 list for the applicable year of decision to code the party characteristics. The complete list of case and party characteristics included in our dataset is provided in Appendix C.

5. *Award Allocation Variables:*

Finally, in cases where multiple patents were asserted, we coded variables to describe the allocation of the ultimate damages award among the patents that were held valid and infringed in order to obtain the patent-specific award amount. This process entailed close examination of the final disposition of each such case, principally relying on verdict forms in jury trials and the order memoranda in bench trials. First, starting from the full set of 221 cases, we separated out those cases in which only a single patent was held valid and infringed. This subset contained 136 cases, and for these we allocated the total award amount to the asserted patent.[63] Of the remaining eighty-five cases in which multiple patents were held infringed, we investigated the opinions and verdicts to identify those cases that explicitly allocated a distinct award amount to each of the infringed patents. There were thirteen such cases, and we assigned the case-specified award to the applicable patents.

For the remaining 72 cases in which multiple patents formed the basis of the award and the court or jury made no explicit, we reviewed the case and docket materials to determine the number of claims that were held infringed for each of the infringed patents. Using these claims counts, we allocated the lump-sum award amount on a *pro rata* basis according to the number of infringed claims in each such patent.[64] Notably, where a verdict or order specified that a

62. Our dataset also includes a flag indicating whether the patent holder is a non-practicing entity, but this study does not analyze that characteristic. A separate study will analyze NPE patents and their characteristics and value.

63. These included both cases in which only a single patent was asserted and cases in which multiple patents were asserted but only a single patent was held valid and infringed.

64. For example, if a case awarded $300,000 for infringement of two patents, *A* and *B*,

particular claim was infringed multiple times, such as by multiple infringing products of the defendant, we counted that claim multiple times for the corresponding patent.[65] Also, sixteen of these cases were directed verdicts for which the case materials did not specify which claims were infringed. For these cases, we allocated the awards *pro rata* based on the number of independent claims of each infringed patent.

B. *Pre-Processing Methodology:*

Following the coding phase, we conducted a number of standard preprocessing steps to facilitate statistical analysis and ensure robustness of the results.[66] We used both Microsoft Excel and R for the preprocessing steps, which included the following:

- Certain of the raw data variables were converted into boolean flags to avoid highly skewed and discontinuous data, such as the numbers of foreign inventors, U.S. child applications, non-patent backward citations and recorded liens, respectively.

- Certain categorical variables were created to represent non-numerical data. For example, IPC codes were categorized into eight groupings based on the first letter industry marker, as shown in Appendix D. Also, SIC codes were grouped into ten categories based on NAICS classification ranges according to the first two digits thereof, as shown in Appendix E.

- After computing patent age and other relevant time durations from the dates we coded (*e.g.*, patent prosecution time was computed as the number of days elapsed between the filing date and issue date), many of these variables were log-transformed to produce normally distributed regressors. For example, prosecution time, litigation length, and the time elapsed prior to assertion (age of patent as of the complaint date) were log-transformed, as were certain count variables where this produced clearer results.

---

and the verdict found 2 claims of patent *A* infringed and 1 claim of patent *B* infringed, then the allocation would be $200,000 for patent *A* and $100,000 for patent *B*. We also note that other allocation methods may be appropriate.

65. One advantage of this approach is internal consistency with cases in which a single patent was infringed—that is, a *pro rata* allocation with a single infringed patent would allocate the entire award to that patent.

66. For example, several preprocessing steps were undertaken to generate normally distributed regressors and eliminate heteroscedasticity thereof in order to avoid unreliable or inflated regression coefficients. These steps, for instance, included log transformations of time and count variables to eliminate over-dispersion, as well as converting certain discontinuous and highly skewed integer variables into categorical factors or boolean flags.

- Finally, the patent-allocated award amount was adjusted for inflation and log-transformed to obtain the dependent regressor for analysis. Specifically, we first applied CPI adjustment to convert each award amount to 2011 dollars. Next, we log-transformed the inflation-adjusted amounts. Such transformation resulted in normally distributed values that allowed us to apply a linear regression model.[67]

C. *Summary Statistics:*

Prior to regression analysis of the full dataset, we generated several key statistics of the distribution of inflation-adjusted patent-specific award amounts. First, Table 1 below shows the mean, median, and standard deviation of the awards distribution, together with skew measures:

Table 1 – Patent Awards Summary Statistics

Statistic	2011 Dollars ($)	2011 Dollars (log)
N	385	385
Mean	$26,172,000	14.58
Median	$2,143,000	14.30
Max	$1,937,849,000	21.38
Variance	1.44e+16	7.46
Std. Deviation	1.20e+8	2.73
Skewness	11.61	-0.34
Kurtosis	172.14	2.80

Next, Figure 1 below plots a histogram of the non-transformed inflation-adjusted distribution, and Table 2 shows the decile amounts:

Figure 1 – Patent Awards Distribution (Histogram)

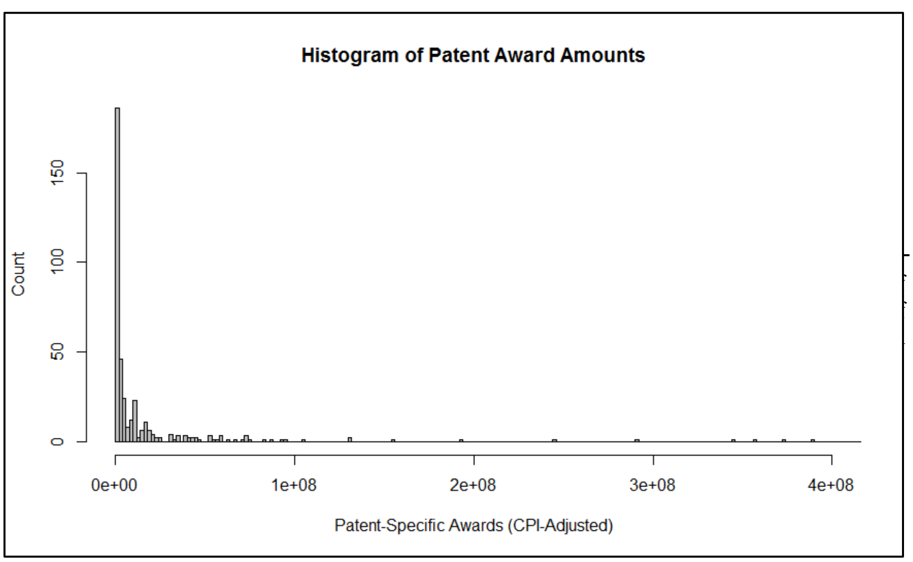


Table 2 – Patent Awards Distribution (Deciles)

Decile	**Max. Award (2011 Dollars)**	**Count**
0%	$644	38
10%	$40,561	38
20%	$137,070	39
30%	$508,010	38
40%	$1,042,990	39
50%	$2,142,857	38
60%	$3,808,688	38
70%	$8,867,199	39
80%	$16,034,170	38
90%	$43,340,160	39
100%	$1,937,849,000	1

We also analyzed the time-dependence of the award values over the six-year period by conducting a targeted regression with the year-of-decision variable and also plotting the respective means and standard deviations of the inflation-adjusted awards in each year. This analysis revealed no significant time trend over the period studied. Table 3 below shows the regression results and Figure 2 shows a box-plot of the awards data by year:[68]

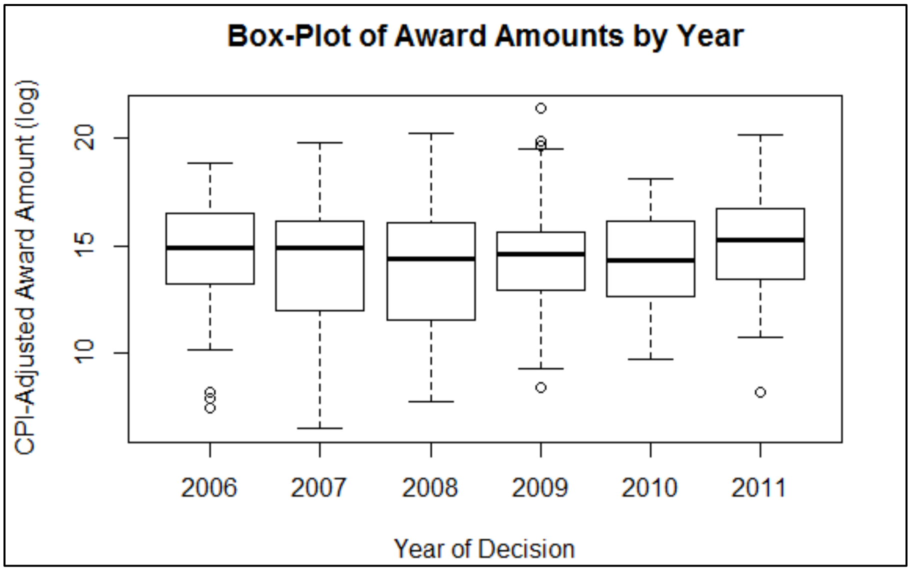
Table 3 – Time Regressions of Award Amounts

Variable	**R^2**	**Adj. R^2**	**Std. Error (df)**	**F Statistic (df1, df2)**	**p-value**
Year of Decision (Integer)	0.0010	-0.0016	2.733 (383)	0.3717 (1, 383)	0.542
Year of Decision (Factor)	-0.0166	-0.0036	2.726 (379)	1.278 (5, 379)	0.272

68. The breakdown of patent-specific award amounts per year are as follows:

2006	2007	2008	2009	2010	2011
52	65	84	76	65	43

Figure 2 – Award Amounts by Year



Finally, we conducted statistical tests to ensure that the log-transformed awards variable was normally distributed, thereby facilitating linear regression modeling. A Kolmogorov-Smirnov test was conducted to compare the log-transformed distribution against a randomly generated normal distribution having the same mean and standard deviation,[69] which found the transformed award amounts to be normally distributed within 5% and 1% degrees of statistical significance.[70]

III. ANALYSIS AND RESULTS:

This section presents the results from our statistical analyses of the complete preprocessed dataset described above. We conducted these analyses using R and its standard functions. Specifically, we ran linear regressions using the normally distributed numerical and boolean or categorical variables described above.[71] We also conducted correlation tests and tested for significant correlations among various patent attributes,[72] using the standard 95% confidence interval unless otherwise noted.

69. Specifically, 10 random normal distribution samples were generated using the mean and standard deviation of the log-transformed awards distribution, and the K-S test results were averaged over these 10 iterations.

70. Specifically, K-S testing over 10 iterations yielded a p-value of 0.481 (not significant) and D-value of 0.0644.

71. Specifically, the `lm` function in R was used for the linear regressions, with the default settings.

72. Specifically, the `cor.test` function in R was used for the correlation tables, with the default Pearson's moment method.

A. *Correlation Analysis*

Our first set of analyses focuses on investigating the characteristics of all patents in the dataset and the relationships between these attributes. Specifically, our dataset contains each of the 385 U.S. patents for which infringement damages were awarded during the relevant time period. Setting aside the question of *how much* value each of these patents has and what drives the differential amounts, we know that each of these patents has *some* enforcement value. That is, each of these patents is a "valuable patent" according to our definition. Accordingly, we first seek to study the characteristics of these patents and learn new insights about the features of "valuable patents" generally.

To do so, we construct a correlation table showing the pairwise correlations among relevant intrinsic and acquired attributes from the dataset. The table is an upper-triangular matrix (only showing values in the upper right-hand diagonal) to eliminate duplication in the transpose cells, and omits identity values along the diagonal.[73] Each cell shows the correlation between the patent attribute in the corresponding row with the attribute in the corresponding column, and we have included an indicator of the degree of significance thereof (*** for significance at the 0.1% level, ** at the 1% level, * for 5% and "." for 10%).

This analysis reveals several important correlations and trends, and the following presents our principal observations grouped according to the categories of intrinsic and acquired attributes described above. The full correlation table for these attributes is provided in Appendix F.

1. *Priority Attributes*

One striking result is the fact that the number of parents from which each patent claims priority is independent of the age-adjusted forward citations of that patent, whereas the number of children applications claiming priority to a patent correlates very significantly with age-adjusted forward citations. Taken together, these results may reflect an important difference between upstream and downstream inventions, whereby upstream inventions that give rise to more follow-on inventions (and therefore more child applications) have a greater technological impact, as represented by the higher number of forward citations received. By contrast, downstream inventions that derive from a longer lineage of parents may have lesser technological impact and fewer forward-citations, or they may be independently important and receive their own citations, leading to no overall correlation between number of parents and citation count.[74] We will further analyze the upstream-downstream distinction

---

73. To read the pairwise correlations from the upper triangular matrix, choose an attribute from the header row and read vertically down to the diagonal, then read right along the corresponding row to the edge.

74. The upstream-downstream distinction has received little attention in scholarship,

in the regression analysis below to determine the differences in enforcement value associated with each type of patent.

We also included in the correlation table the patent filing year, which exhibits several correlations one would expect and therefore provides a sanity check for the data.[75] For example, the time to assert (age at complaint) and numbers of assignments and liens are significantly negatively correlated with filing year, due to time truncation effects for younger patents.

The number of parent applications is significantly correlated with the number of inventors,[76] length of the written description, and number of figures, each of which would be expected for continuation and other child applications that are likely to claim additional subject matter and have greater specificity than more primary upstream applications.

Interestingly, the number of parents is also significantly positively correlated with the number of Office Action rejections received during prosecution. This could reflect the process of refining claims of the child applications to distinguish it from their parents. Alternatively, child applications with a greater number of parents could be facing a greater number of rejections because applicants are trying to test the limits of their priority and claim scope with new claims that arguably incorporate new matter or are not fully enabled by the earlier-filed specifications.[77]

2. *Inventor Characteristics*

There are also several other interesting correlations involving the number of inventors. For instance, patents with more inventors tend to have more claims, longer written descriptions, and more figures. These correlations likely reflect the added subject matter contributed by each inventor.[78] Also, patents with more inventors tend to have more child applications, which could also reflect additional subject matter contained in the original specification that can be claimed by future child applications.[79] Finally, patents with more inventors

---

although there are some notable exceptions. *See*, *e.g*., Dmitry Karshtedt, *The Completeness Requirement in Patent Law*, 56 B.C. L. REV. (*forthcoming* 2015) (defining "upstream inventions" as patents on research tools that could block subsequent inventive activity).

75. The priority year and publication status of each patent are omitted from the correlation table as they are strongly correlated with filing year and number of parents, respectively, and therefore exhibit similar pairwise results.

76. Similarly, for child applications with more parents, given that all or a subset of the child's named inventors will also be named inventors of the parent applications (and therefore have more hits on average).

77. Alternatively, applicants may file continuation claims that are not substantively different than the parent applications, which should also generate more rejections.

78. Inventorship requires contribution to the conception of the claimed invention, not merely reduction to practice, and therefore more inventors are likely to correspond to more claimed subject matter. *See Fiers v. Revel*, 984 F.2d 1164, 1169 (Fed. Cir. 1993).

79. Also, as noted above, the number of inventors is significantly positively correlated with the number of parents (i.e., later-stage child applications are more likely to have more inventors), which is expected based on the rules of inventorship and continuation practice.

tend to receive more forward citations, which may indicate greater technological impact (although this may also in part reflect citations by future patents of these inventors, which collectively may be greater in number than patents produced by fewer inventors).

3. *Recorded Transfers and Liens*

Another interesting result is the independence of recorded assignments and liens from many intrinsic patent attributes. Specifically, the numbers of recorded assignments and liens are not significantly correlated with priority attributes, structural features, or inventor characteristics.[80] Notably, to a certain extent one would expect a degree of independence of recorded transfers and liens from intrinsic patent characteristics, given the blanket approach typically taken to recordation practices. That is, where a business is being sold or bank financing is being provided, the buyer/lender will typically record against all acquired/pledged assets given the low costs associated with doing so (at least for domestic intellectual property), rather than select only the most valuable assets for recordation.

Yet, we do find that forward citations are significantly correlated with recorded assignments. This is consistent with recent findings of Serrano, who conducted a large-scale empirical study of transferred patents and found evidence that patents that have higher private valuable to their owners (as proxied by forward citations) are more likely to be transferred.[81] Yet, it also suggests some ambiguity in the types of value indicated by forward citations—specifically, as we find below, transfer value is often inverse to enforcement value and therefore if forward citation counts indicate both types of value then a higher number of forward citation alone does not necessarily suggest a high expected infringement award.

4. *Forward Citations*

Forward citations are one of the most strongly correlated patent attributes in the dataset, and they exhibit significant correlations with nearly every other variable. For example, age-adjusted forward citations are significantly positively correlated with the number of inventors, prosecution time and backward citations (both applicant and examiner citations), structural features (number of claims, length of written description and number of figures), as well as number of children (but not number of parents).

As noted above, these results further suggest that forward citations in bulk

---

80. Although the number of assignments is significantly correlated with the number of children, and significantly negatively correlated with the filing year (or positively correlated with the age at trial/time-to-assertion), these results are likely driven by age truncation effects.

81. Serrano, Characteristics of Patent Litigation: A Window on Competition, *supra* note 55, at 686-708.

may be somewhat imprecise identifiers of particularly valuable patents. By contrast, more refined citation metrics could yield greater clarity. For example, in measuring relationships with stock market value, Hall, Jaffe and Trajtenberg distinguish between self-citations (from subsequent patents owned by the same applicant as the cited patent) and citations from third party patents.[82] Also, Lanjouw and Schankerman use the number of citations per claim and sophisticated measures of technology diffusion and technology area concentration to study the determinants of patent litigation.[83] More recently, Galasso and Schankerman also employ refined citation metrics to study the impact of patent invalidation on subsequent technology development (*e.g.*, new third party citations received after invalidation).[84] In future research we intend to parse citation data in closer detail to investigate which specific types of citations are most strongly associated with enforcement value.

B. *Regression Models:*

Next, we conducted regression analysis to study the relationship between various patent attributes and enforcement value. We first attempted to maximize the degree of fit of the model by incrementally adding variables and measuring the increased explanatory power they provide versus reductions in degrees of freedom. Table 4 below shows the degree of fit of various iterations of this process as well as the final resulting model:

82. Hall et al., *supra* note 26.

83. Lanjouw and Schankerman, *supra* note 15.

84. Alberto Galasso, & Mark Schankerman, *Patents and Cumulative Innovation: Causal Evidence from the Courts*, 130 Q.J. Econ. 317 (2014).

Table 4 – Fitting the Regression Model

Model	R^2	**Adj.** R^2	**Std. Error (df)**	**F Statistic (df1, df2)**	**p-val**	**Sig. of Change (ANOVA)**
(1) Baseline Model	0.5954	0.5470	1.838 (343)	12.31 (41, 343)	$< 2.2^{-16}$	N/A
(2) w/ Upstream & Downstream	0.6236	0.5723	1.786 (338)	12.17 (46, 338)	$< 2.2^{-16}$	***
(3) w/ Forward Citations	0.6315	0.5802	1.769 (337)	12.29 (47, 337)	$< 2.2^{-16}$	**
(4) w/ Assign. & Liens	0.6459	0.5941	1.740 (335)	12.47 (49, 335)	$< 2.2^{-16}$	**
(5) w/ Prosec. History	0.6729	0.6182	1.687 (329)	12.31 (55, 329)	$< 2.2^{-16}$	***
(6) w/ Inventor Characteristics	0.6807	0.6262	1.670 (328)	12.49 (56, 328)	$< 2.2^{-16}$	**
(7) w/ Structural Features	0.6931	0.6374	1.644 (325)	12.44 (59, 325)	$< 2.2^{-16}$	**
(8) All variables	0.6972	0.6411	1.636 (324)	12.43 (60, 324)	$< 2.2^{-16}$	*

Full regression results on file with the author.

Specifically, we started with a small baseline subset of independent variables and iteratively added new variables to measure their impact on the model. The significance testing shown in the last column of Table 4 are analysis of variance (ANOVA) measures based on the immediately preceding model, showing (loosely speaking) the extent to which each new model has significantly better fit than the preceding one.

The final regression reveals a considerably high degree of fit. Specifically, Model (8) includes all categories of pre-processed variables in the regression and yields an R^2 value of approximately 0.70. Loosely speaking, this means that approximately 70% of the variation in award value is explained by the variables, which is quite high for this type of statistical analysis.[85] This degree of fit further suggests that the majority of systematic factors influencing award value are represented by the modeled variables, particularly since one would expect a fairly high amount of idiosyncratic variation in award value that would limit the theoretical maximum degree of fit of even a perfect model. For example, specific features of the individual litigations, infringing products and parties in each case are also likely to influence award value, and jury or judge

85. This result is also consistent with previous analyses of patent infringement awards (including studies by the author), which provide further support. *See* Mazzeo et al., *Explaining the "Unpredictable"*, *supra* note 5, at 66-67.

idiosyncrasies may also in higher or lower award amounts than expected. Such factors are not susceptible to modeling and limit the best possible fit.

Finally, we provide in Appendix G plots of the residuals of the final model, showing that the errors are generally normally distributed. This suggests that there are no strong systematic trends missing from the model.

C. *Specific Variable Analyses*

Our final set of analyses focuses on the specific relationships between enforcement value and certain key patent, case and party variables, based on significance testing of their specific regression coefficients. We used the regression coefficients from Model (8), shown in Appendix H, as well as certain targeted analyses described below.

1. *Case and Party Characteristics*

Some of the most significant factors influencing patent enforcement value are the characteristics of the parties and features of the litigations in each case, independent of attributes of the asserted patents themselves. This is not surprising. For example, whether a case is decided by judge or jury is expected to significantly influence award amount, in large part because high-stakes litigations are most commonly tried by juries (due to litigant selection effects). Similarly, defendants in the Fortune 500 are more likely to face larger infringement verdicts because their revenue base for measuring damages is likely to be larger than that of a smaller infringer.[86] Also, parties have greater incentives to litigate more aggressively and spend more on their claims and defenses when the stakes are higher, as reflected by the significant correlation between litigation length and award value.

However, one key result is the significant positive coefficient for lost profit awards,[87] whereas reasonable royalty awards are not significant.[88] The fact that

86. These results are consistent with previous published analyses by the author. *See* Mazzeo et al., *Do NPEs Matter?*, *supra* note 5, at 3. However, more recent data indicates that the median NPE award is significantly higher than the median practicing entity award. 2014 PwC Study, *supra* note 40, at 19.

87. This is consistent with previous findings. *See* Mazzeo et al., *Explaining the "Unpredictable"*, *supra* note 39, at 68. However, the majority of awards (209 in total) are reasonable royalty awards, which is consistent with PwC's reports. *See* 2014 PwC Study, *supra* note 40, at 9.

88. We note that previous work found jury awards of reasonable royalties to be significantly positively correlated with case-aggregated (as opposed to patent-specific) award amounts. There are several possible explanations for the difference here. For example, plaintiffs who claim reasonable royalties may be more likely to assert multiple patents (and therefore be more likely to obtain a smaller award for each patent) than plaintiffs who claim lost profits, particularly if plaintiffs seeking reasonable royalties are not suing their direct competitors where it may be more likely that a single core patent covers the plaintiff's technology and is infringed by the defendant's accused product. The data provide some support for this hypothesis, specifically that reasonable royalties were the basis of damages

lost profits tend to be higher indicates that patent holders who practice their inventions are likely to receive larger infringement verdicts. Importantly, lost profits are available only where the patent holder is practicing the invention in competition with the accused infringer.[89] Thus, this result suggests that *commercialized patents* tend to have higher enforcement values.[90]

It should also be noted that the patent holder's Fortune 500 status is significantly positively related with award amount, whereas the patent holder's small entity status is significantly negatively related. These factors indicate that the size of the patent holder strongly influences award value. To the extent patent holder size corresponds to the extent of commercialization value derived from the technology at issue (which is a reasonable assumption, particularly for lost profit awards which are based on the profit the patent holder would have earned from its own commercialization activities but for the infringement), this further points to the link between commercialization and enforcement value.

2. *Priority Claims and Child Applications*

Based on our observations from the correlation analysis above, we further seek to identify whether upstream patents (with fewer parents and more children applications) have significantly higher or lower enforcement value than downstream patents (with more parents and fewer children). We find strong evidence that upstream technology patents tend to have higher enforcement values than downstream patents. Patents that give rise to child applications tend to have significantly higher enforcement award values overall, and patents that are children of other patent applications (i.e., continuations, divisionals or continuations-in-part) tend to have significantly lower enforcement values.

These results emphasize the distinction between upstream and downstream patents and the strong impact this has on expected enforcement value. Importantly, each category is significantly distinguishable from the population of other patents on the whole based on the significance of their respective

in 56% of cases in which multiple patents were held infringed, whereas they were the basis of damages in only 46% of cases in which a single patent was held infringed.

89. The lost profits calculus generally requires such competition. *See* Panduit Corp. v. Stahlin Bros. FibreWorks, 575 F.2d 1152, 1157 (6th Cir. 1978).

90. Furthermore, this result is contrary to many widespread concerns about non-practicing entities. Rather, the highest enforcement values are recovered by parties who enforce their patents as a means of protecting their own product-driven revenues, whereas parties who do not practice the inventions claimed by their asserted patents are likely to receive lower awards. In an economy of high enforcement costs, this result also implies that the return on investment earned from enforcing patents tends to be lower for non-practicing entities, except to the extent they are able to reduce their own enforcement costs relative to other patent holders. Common NPE practices reflect this need to streamline enforcement: e.g., some of the most successful NPEs often utilize contingent-fee legal services and gravitate to "rocket-docket" jurisdictions that may help to lower their costs and secure payment more efficiently.

coefficients.[91] Furthermore, the significant differences between the intrinsic characteristics of upstream versus downstream patents identified in our correlation analysis above provides additional support to the categorical distinction we observe here.[92]

Notably, child applications are filed by the applicant with respect to their own prior parent applications,[93] and therefore they are most properly viewed as indicators of upstream or downstream inventions relative to the applicant's *own* technology.[94] This is consistent with the findings of Hall et al. that self-citations are significant indicators of value, as self-citations similarly arise from future patents of the same applicant.[95] Additionally, this also comports with the findings of Abrams et al. that patent holders may protect their market share by strategically filing downstream child patents to crowd out competitors. One might expect greater enforcement value to accrue to upstream patents as a result of such strategic activities.[96]

However, it is also reasonable to expect more continuations to be filed for upstream inventions that are more valuable to begin with, as the original applicant mines this novel technological resource and further refines and improves the original technology. For example, theories of cumulative innovation by Kitch and others suggest that one main social benefits of patent protection is to facilitate subsequent development of early-stage inventions and

---

91. Also, as a robustness test we checked that there were reasonable populations of patents that were not clearly upstream or downstream. Specifically, 42% of patents had *both* at least one parent and child application, and 13% of patents had *no* parent or child applications.

92. Notably, the upstream-downstream distinction should not be confused with the difference between platform versus end-user technologies. For example, the foundational architecture of mobile technologies or other standard-essential inventions may be considered platform technologies, which give rise to improvement technologies created by *third parties*. As a result, platform technologies are likely to be widely licensed, such as via FRAND or other non-exclusive cross-licensing terms, which would suggest that their *enforcement* value may be lower than their aggregate *transaction* value. Further distinction between these various categories is a topic for future research.

93. One semantic exception is when an examiner issues a restriction requirement during prosecution, which requires the applicant to move certain claims to a divisional application (thereby creating another child application) or cancel them. Yet, even in this circumstance the applicant has the option to elect whether to continue prosecuting the divided claims.

94. The applicant-specific definition makes sense, particularly when innovation is viewed as an iterative process of intra-firm improvement with limited cross-pollination between competing technologies. *See*, *e.g.*, Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & Econ. 265 (1977). Also, these results are consistent with the commercialization-effect observed in the lost profits variable and other regressors—a patent holder that improves upon its own technology is more likely to practice it as well, and the data indicates that these types of patent holders receive the highest enforcement value in infringement suits.

95. Hall et al., *supra* note 26 at 4.

96. Abrams et al., *supra* note 33 at 7.

promote downstream technological advancement.[97]

3. *Structural Features*

Notably, structural features of the patents studied are generally not significant. For example, we tested the number of figures (as well as a flag indicating whether a particular patent is in the top quartile with respect to number of figures), presence of formulas or data tables, claim length and number of dependent claims per independent claim, and we found that these are not individually significant. Also, adding or omitting them in the aggregate does not significantly affect the fit of the overall model.

However, one notable exception is the length of the written description, which is positive and significant at the 1% level. We further tested whether this result was influenced by variations in typical written description lengths in different technology sectors—for example, if high-tech patents in IPC G or H are likely to have longer written descriptions, then the significance we observe here could simply be driven by technology effect (as noted below). Notably, we do find substantial variation between average WD-length in each IPC category, with an inter-quartile increase of nearly 100% (*i.e.*, nearly double) from the first quartile to the third quartile, and IPC A and H registering as having the highest average WD lengths. However, excluding only the WD length variable significantly reduces the model's fit, which suggests that its explanatory effect is not entirely related to IPC category. Also, when we replace the raw WD length with IPC-adjusted metrics, such as mean-removed WD length (where the IPC-specific mean WD length is removed) and WD length percentile measures (where the IPC-specific WD length distribution is used to determine the percentile), we still find these metrics to be positive and significant.

Optimistically, the positive relationship between written description length and enforcement value could indicate that patents with greater disclosure yield larger rewards to their owners. If true, this could support the disclosure function policy justification for granting patent rights. However, we do not have sufficient evidence in this study to infer such a result, and we leave this topic for future research.

4. *Prosecution History and Backward Citations*

We also find certain significance of the prosecution duration and number of Office Action rejections, but these results are ambiguous. In particular, extremely high prosecution lengths are negative and significant, as measured by a boolean flag indicating whether the prosecution was in the top 75 percent of all patents in the sample. The number of Office Action rejections is positive and significant only when this prosecution length flag is added, which could suggest that these effects are cancelling each other out to a certain extent.

97. Kitch, *supra* note 94.

Also, the total number of backward citations is not significant individually. Although it is positive and significant in certain models that also include a citations per claim variable, the latter has a negative coefficient which suggests the effects may be cancelling each other.

5. *Inventor Characteristics*

Inventor characteristics generally do not exhibit strong correlations with enforcement value, including the total number of inventors and the number of U.S. inventors, as well as a flag indicating whether there are many inventors relative to the average. However, the presence of a foreign inventor is positive and significant at the 1% level. We expect this relates to the size of the patent holder, whereby larger firms are more likely to have foreign R&D facilities and also foreign firms that are selling and enforcing patented products in the U.S. market are more likely to have larger scale.

6. *Forward Citations*

As expected based on prior studies, forward citations are significantly positively correlated with enforcement value. We use the natural logarithm of citations received from later patents that had been filed as of the complaint date, and we find this measure to be strongly positively and significantly associated with patent enforcement value (at the 0.1% level). We also find significance of base 10 (not log-transformed) forward citation counts, although these results are somewhat weaker and only apparent in certain model specifications.

As discussed above, we expect some degree of age dependency and truncation effects given that several of the patents in our sample were still relatively young at the time they were asserted. Accordingly, we also measured the citations received from all patents issued as of Dec. 2014, which allows us to capture a much longer portion (if not all) of the expected high citation window of these patents. Replacing this variable in the analysis does not significantly change the fit of the model or overall results, and we find this measure of citations to similarly be positive and significant at the 0.1% level.

Notably, we also observe that when we use the Dec. 2014 citation counts, the coefficient of patent age becomes significant (at the 10% level), which suggests that, although weak, patent age has a positive relationship with enforcement value. This result makes sense given the longer timeframe during which damages may accrue for older patents. However, the effect is weak, and age interactions with forward citations may mask it in certain specifications.

Notably, as mentioned above, our correlation analysis reveals that forward citations are also significantly positively correlated with many other intrinsic and acquired attributes, which themselves are not significant indicators of patent value based on our regression results. Accordingly, forward citations may face Type I errors (false positives) as predictors of patent enforcement value, in that a high number of forward citations may be associated with other characteristics that do not necessarily translate into greater enforcement value. Nonetheless, particularly when viewed in conjunction with other factors in the context of a robust model, forward citations do exhibit strong associations with enforcement value and thus have utility in prediction and analytics applications.

7. *Forward Citation-to-Value Relationship*

Motivated by the recent findings of Abrams et al. that forward citation counts do not monotonically increase with patent value, we further investigate the citation-to-value relationship in the context of enforcement awards. Specifically, Abrams et al. find that citations follow an inverted U-shaped curve relative to value, whereby higher-value "strategic" patents exhibit a decreasing citation-to-value relationship.[98] First, we test whether the citation-to-value relationship for our dataset exhibits an inverted-U shape similar to that reported by Abrams et al. We bucketed patent value into percentiles and calculated the average number of forward citations (using the December 2014 counts) received by patents in each bucket. We also plot citations received on the y-axis versus non-bucketed values on a log scale on the x-axis. Appendix I shows each graph. Interestingly, we do not observe an inverted-U curve or decreasing citation-to-value relationship at the high-value range, although we do observe a tapering off for the most valuable patents that suggests a non-monotonically increasing relationship.

To investigate further, we regress forward citations against absolute (base 10) enforcement value and the square of enforcement value to see if we can detect a significant negative quadratic term consistent with their findings. As reported in Appendix J, the coefficient of value squared is negative and significant, although it is much weaker in magnitude than the positive linear value term.[99] This is consistent with Abrams et al.'s observation of a negative quadratic influence, although they observe a much stronger effect that dominates at the high-value range.[100] By contrast, when we add trend lines to the figures in Appendix I to show both a purely linear relationship (shown in red) versus a linear-plus-negative quadratic relationship (shown in blue) using our regression coefficients, we find that the negative quadratic effect is quite minor even at the high end. Overall, the relationship between citations and value we observe is always increasing, and the values required for the quadratic term to dominate and the trend to begin to slope downwards would be extremely high.[101]

These results provide some perspective on the differences between

98. Abrams et al., *supra* note 33, at 7.

99. These results are consistent when we add certain party and patent characteristics, such as patent holder size, technology sector and patent age, consistent with Abrams et al.'s analysis. *Id*.

100. *Id*. at 28.

101. The local maximum implies a patent award value of $2 billion in 2008 dollars.

licensing value and enforcement value. Particularly given the selection differences between Abrams et al.'s dataset and ours. They use values of voluntary licensing revenues received by a large NPE in the high-tech sector, whereby revenues are generated by widely licensing the patents on a non-exclusive basis to multiple companies simultaneously. By contrast, our data is derived from individual enforcements in which the patent holder and infringer do not agree to a license, and the more valuable patents are likely to be those that were not previously licensed but were practiced exclusively by their owners. Furthermore, whereas commercialization value and ongoing technological development (upstream patent value) are principal drivers of our data, neither of these effects is present for NPE-owned patents that in large part have been acquired (directly or indirectly) from the original applicants.

Nonetheless, consistent with their results, we also observe a negative quadratic influence, which suggests that citations do not monotonically increase with value. This provides additional impetus to investigate the citation-to-value relationship more closely, which we plan to study in future work.

8. *Technology Field*

We also analyzed the IPC technology classifications to determine if certain technology fields are significantly associated with higher or lower enforcement value. We find that specific IPC categories exhibit significant relationships, with the high-tech sector (IPC G for Physics and H for Electricity) in particular showing a strong positive association. The bio-pharma sector (IPC category A for Human Necessities) is also positive and significant, although to a lesser extent. These results reflect the fact that patents are highly important assets in the high-tech and bio-pharma sectors. However, the selection of the data may also influence these results, in that widespread licensing practices and patent densities in the high-tech sector suggests that the cases that go to trial (rather than settle) may be more likely to be extremely contentious and have even higher stakes than litigated cases in other industries.

9. *Recorded Transfers and Liens*

Another striking result is the strong negative coefficient for the number of recorded assignments. As posited above, entities that practice and derive commercialization value from their patents may be less likely to transfer them, which is further supported by our finding that commercialization value is a strong contributor to enforcement value.[102] Moreover, we find that patents that are transferred more often (and therefore presumably have higher transaction

102. Also, non-practicing entity litigation may contribute, in that NPEs typically acquire their patents from other entities (and therefore increase the number of assignments), and NPEs tend to receive lower awards on average. *See* Mazzeo et al., *Do NPEs Matter?*, *supra* note 5, at 3.

value) tend to have even lower enforcement value, as measured by the specific boolean flag indicating whether the patent has been assigned many times (*i.e.*, the number of recorded assignments is in top 75% quartile).

By contrast, the boolean flag for recorded liens is not significant in either regression model. As mentioned above, this could reflect the blanket approach taken to collateralization and perfection in many financing arrangements. Also, the small number of recorded liens in the dataset overall could reflect lender restrictions on enforcing patents that have been pledged as collateral.[103]

IV. INTERPRETATIONS AND CONCLUSIONS:

The analysis above reveals three distinct categories of "valuable patents" and the characteristics of each; additionally we observe fundamental distinctions between patent enforcement value and transaction value. The following summarizes our principal findings:

A. *Commercialization Value*

We find that one of the most significant indicators of patent enforcement value is the extent to which a patent is practiced by its owner. Lost profit damages, which are only available to practicing patent holders, are strong predictors of award amount, and the size of the plaintiff, which one would expect to reasonably correspond to the size of the revenue base associated with the technology at issue, is also significantly associated with award value. Also, lower enforcement awards are associated with patents that have been transferred, which further highlights the impact of commercialization by the patent holder. Specifically, a non-commercialized patent may yield greater value as an asset to be traded than as exclusive rights to be enforced, whereas a patent covering valuable commercialized technology is more likely to be retained to protect its owner's competitive domain.

B. *Upstream Technology Value*

We also observe clear evidence of upstream technology value as another category of patent enforcement value. Specifically, patents that give rise to follow-on child applications are likely to result in higher awards, both on average and in particular relative to downstream patents that claim priority from parent applications.

C. *Citation Value*

We also find that forward citations are an important indicator of patent enforcement value. This is consistent with many other studies, and common

103. Only 49 (approx. 13%) of patents in the dataset had recorded liens.

analytics practices, that treat forward citations as a general proxy for patent value. Yet, we further find that forward citations are highly correlated with a host of other patent attributes, some of which are associated with high enforcement value and others with low enforcement value, which suggest that forward citations may simultaneously code for multiple different types of patents and sources of value. Moreover, we find that the citation-to-value relationship is not strictly linearly increasing, and particularly in the high-value range there are negative influences that reduce the observed association. We plan to investigate the value predicting capacity of more refined citation-based metrics in future research.

D. *Transfer Value*

Finally, we find evidence of key distinctions between patent enforcement value and transaction value. Specifically, patents that have been assigned tend to have lower enforcement values, and patents with many recorded assignments are further associated with lower awards. Furthermore, the number of recorded assignments and liens are independent of most intrinsic patent characteristics. By contrast, we find that assignments and liens are significantly correlated with forward citations, which suggests that citations may be associated with both enforcement value and transaction value.

Accordingly, based on our analysis of patent enforcement value, we find that "valuable patents" are (1) commercialized patents, (2) upstream patents, and (3) with certain caveats, forward-cited patents. Current theory and practice use a definition of "valuable patents" that is derived from proxy-based studies of patent value, but there are substantial limitations to this approach. Instead, by studying value directly, we derive new findings about factors that define, and in turn may be used to predict, the enforcement value of U.S. patents.

APPENDICES:

APPENDIX A - INTRINSIC PATENT ATTRIBUTES:

Variable	**Description**	**Coding Source**
us_patent_no	Patent number	Page 1 of Patent
us_pub_no	Publication number	Page 1 of Patent
file_date	Filing date	Page 1 of Patent
published?	TRUE if the application was published (has a publication number), FALSE if not.	Page 1 of Patent/PAIR
pub_date	Publication date	Page 1 of Patent
priority_date	The priority date shown in Google Patents	Google Patents
issue_date	Issue date	Page 1 of Patent
IPC	IPC Code (e.g., G04FFalse)	Page 1 of Patent
utility?	TRUE if this is a utility patent (TRUE *unless* patent number starts with "D" or "PP")	Page 1 of Patent
design?	TRUE if this is a design patent (patent number starts with "D")	Page 1 of Patent
plant?	TRUE if this is a design patent (patent number starts with "PP")	Page 1 of Patent
cont?	TRUE if this is a continuation of another patent or patent application	Page 1 of Patent
div?	TRUE if this is a divisional of another patent or patent application	Page 1 of Patent
cont_in_part?	TRUE if this is a continuation-in-part of another patent or patent application	Page 1 of Patent
num_parents	Number of parent patents or patent applications referenced on page 1 / PAIR	Page 1 of Patent / PAIR > Continuity Data
PCT?	TRUE if this is a PCT patent	Page 1 of Patent (PCT number)
PCT_number	PCT number assigned	Page 1 of Patent
term_disclaimer?	TRUE if there is a terminal disclaimer	Page 1 of Patent (asterisk next to patent number / notice)
term_adjustment?	TRUE if there is a patent term adjustment	Page 1 of Patent (notice)
num_US_inventors	Number of inventors with US domiciles	Page 1 of Patent

Variable	Description	Coding Source
num_foreign_inventors	Number of inventors with foreign domiciles	Page 1 of Patent
num_inventors	Sum of num_US_inventors and num_foreign_inventors	Calculated
entity_status	Entity status as shown in PAIR (e.g., "Small", "Undiscounted")	PAIR (http://portal.uspto.gov/pair/PublicPair)
named_applicant?	TRUE if there is a named applicant in PAIR; FALSE otherwise.	PAIR (http://portal.uspto.gov/pair/PublicPair)
bkd_cit_applicant	Number of citations to other patents/apps that were made by the applicant.	Page 1 of Patent
bkd_cit_examiner	Number of citations to other patents/apps that were made by the examiner (*).	Page 1 of Patent
bkd_cit_tot	Sum of bkd_cit_applicant and bkd_cit_examiner	Calculated
non_pat_cites?	TRUE if the patent cites non-patent documents.	Page 1 of Patent
non_pat_cites_num	Number of non-patent documents cited in the patent.	Page 1 of Patent
US_children?	TRUE if patent has any patents or applications claiming priority to it	PAIR > Continuity Data (http://portal.uspto.gov/pair/PublicPair)
num_US_children	Number of patents or patent applications claiming priority to this patent	PAIR > Continuity Data (http://portal.uspto.gov/pair/PublicPair)
num_OA_rejections	Number of Office Action Rejections (Final or Non-Final) during prosecution.	PAIR > Transaction History
reexam_reissue?	TRUE if this is a reexamination patent (number starts with RE OR reexam cert)	Patent / Patent Number
re_number?	TRUE if the patent number starts with RE	Patent Number
reexam_cert?	TRUE if there is a Reexamination Certificate appended to the patent	Patent
cert_correction?	TRUE if there is a Correction Certificate appended to the patent	Patent
num_claims_tot	Total number of claims (adjusted for re-exam/correction, if applicable)	Patent (count)
num_indep_cl	Number of independent claims (adjusted for re-exam/correction, if applicable)	Patent (count)
num_dep_cl	Number of dependent claims (adjusted for re-exam/correction, if applicable)	Patent (count)
num_cols_WD	Number of columns in the written description portion of the patent (whole cols)	Patent (count)
num_cols_cl	Number of columns in the claims portion of the patent (whole cols, at least 1)	Patent (count)
num_figures	Number of figures in the patent	Patent (count)

Variable	Description	Coding Source
formulas?	TRUE if the patent contains mathematical or scientific formulas; FALSE otherwise.	Patent
num_formulas	Number of mathematical or scientific formulas in the patent.	Patent (count)
data_tables?	TRUE if the patent contains tables of data; FALSE otherwise.	Patent
num_data_tables	Number of tables of data in the patent.	Patent (count)

APPENDIX B – ACQUIRED PATENT ATTRIBUTES:

Variable	Description	Coding Source
fwd_cites_prev	Total number of forward citations prior *to* the complaint date.	USPTO "Referenced By"
num_assignments	Number of assignments prior to the complaint date.	Goog Pats / PAIR
num_liens	Number of liens prior to the complaint date.	Goog Pats / PAIR

APPENDIX C – CASE AND PARTY CHARACTERISTICS:

Variable	Description	Coding Source
num_P	Number of plaintiffs in the final decision awarding damages.	Complaint/Docket
num_P	Number of defendants in the final decision awarding damages.	Complaint/Docket
p_ind_sic_4	Industry SIC Code (4-digit) of the Plaintiff.	Mergent: mergentonline.com
d_ind_sic_4	Industry SIC Code (4-digit) of the Defendant	Mergent: mergentonline.com
p_US_F500_csyr_B	TRUE/FALSE if Plaintiff was listed on the US Fortune 500 *for the year of decision (2006-2011)*.	US Fortune 500: http://fortune.com/fortune500/
p_US_F500_csyr_rank	Rank of Plaintiff on the US Fortune 500 *for the year of decision (2006-2011)*, if p_F500_caseyr_B is TRUE.	US Fortune 500: http://fortune.com/fortune500/
d_US_F500_csyr_B	TRUE/FALSE if Defendant was listed on the US Fortune 500 *for the year of decision (2006-2011)*.	US Fortune 500: http://fortune.com/fortune500/
d_US_F500_csyr_rank	Rank of Defendant on the US Fortune 500 *for the year of decision (2006-2011)*, if p_F500_caseyr_B is TRUE.	US Fortune 500: http://fortune.com/fortune500/

APPENDIX D – IPC CODE CATEGORIES:

A: Human Necessities
B: Performing Operations, Transporting
C: Chemistry, Metallurgy
D: Textiles, Paper
E: Fixed Constructions
F: Mechanical Engineering, Lighting, Heating, Weapons
G: Physics
H: Electricity

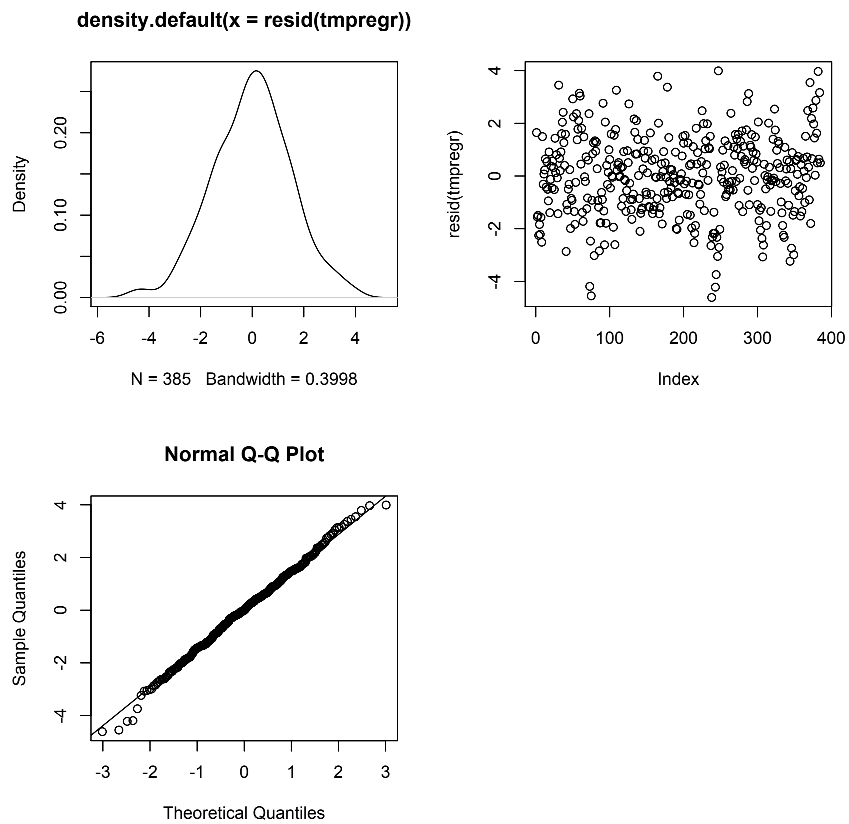
APPENDIX E – SIC CODE GROUPINGS:

01-09: Agriculture, Forestry & Fishing
10-14: Mining
15-17: Construction
20-39: Manufacturing
40-49: Transportation, Communications, Electric, Gas & Sanitary Services
50-51: Wholesale Trade
52-59: Retail Trade
60-67: Finance, Insurance & Real Estate
70-89: Services
91-99: Public Administration

APPENDIX F – CORRELATION TABLE:

	prosec. time	# OA rej.	# par.	# inv.	bkd. cit. app.	bkd. cit. ex.	# child.	# claims	WD length	# figs.	time to assert	fwd. cit.	# assg.	# liens
file year	0.03	-0.03	0.26 ***	-0.03	0.31 ***	-0.04	0.07	0.12 *	0.08	0.08	-0.56 ***	-0.06	-0.33 ***	-0.11 *
prosec. time		0.44 ***	0.07	0.14 **	0.04	0.05	0.12 *	0.13 **	0.17 ***	0.11 *	-0.18 ***	0.11 *	0.07	0.07
# OA rej.			0.14 **	0.19 ***	0.04	0.16 **	0.1 *	0.03	0.07	0.07	-0.09 .	-0.01	0.05	0.11 *
# par.				0.15 **	0.36 ***	0.03	0.1 .	0.23 ***	0.29 ***	0.27 ***	-0.27 ***	-0.01	-0.11 *	-0.07
# inv.					0.09 .	0.05	0.2 ***	0.16 **	0.31 ***	0.34 ***	-0.07	0.22 ***	-0.01	0.01
bkd. cit. app.						-0.12 *	0.13 **	0.22 ***	0.12 *	0.18 ***	-0.23 ***	0.13 *	-0.11 *	-0.04
bkd. cit. ex.							0.01	-0.04	-0.05	-0.04	0.08	0.12 *	0.06	-0.03
# child.								0.18 ***	0.24 ***	0.2 ***	-0.01	0.34 ***	0.14 **	0.01
# claim.									0.33 ***	0.26 ***	-0.05	0.28 ***	0.07	0.11 *
WD length										0.71 ***	-0.04	0.1 *	-0.05	-0.04
# figs.											-0.05	0.12 *	-0.05	0.04
time to assert												0.34 ***	0.26 ***	0.11 *
fwd. cit.													0.33 ***	0.16 **
# assg.														0.18 ***

APPENDIX G – MODEL RESIDUALS:



APPENDIX H – REGRESSION RESULTS:

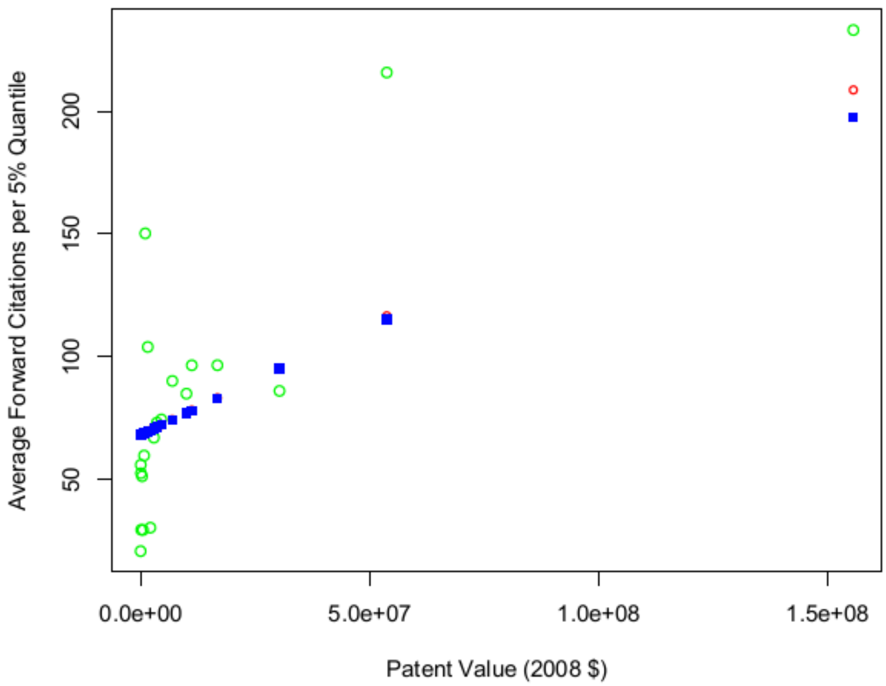
Variable	Coeff.	Std.Err.	t-val	Pr(>|t|)	Signif.
Jury Trial	1.52	0.29	5.19	0.00	***
F500 Patent Holder	0.84	0.31	2.71	0.01	**
F500 Defendant	1.13	0.29	3.90	0.00	***
Lost Profits	1.15	0.27	4.30	0.00	***
Reasonable Royalties	0.20	0.21	0.96	0.34	
NPE (individual)	0.54	0.44	1.22	0.22	
NPE (company)	0.50	0.33	1.51	0.13	
NPE (university)	3.45	1.40	2.47	0.01	*
Small Entity	-0.92	0.25	-3.70	0.00	***
IPC A	1.27	0.71	1.77	0.08	.
IPC G	1.66	0.67	2.49	0.01	*
IPC H	1.84	0.69	2.66	0.01	**
Patent Age	0.00	0.00	0.93	0.35	
Parent?	0.87	0.27	3.26	0.00	**
Child?	-0.74	0.25	-2.96	0.00	**
Forward Citations	0.38	0.10	4.01	0.00	***
Assigned?	-0.61	0.25	-2.41	0.02	*
Many Assignments?	-0.62	0.27	-2.32	0.02	*
Foreign Inventor?	0.94	0.29	3.27	0.00	**
Long Prosecution?	-0.68	0.26	-2.59	0.01	*
# Office Actions	0.18	0.09	2.02	0.04	*
WD Length	0.01	0.00	3.11	0.00	**

Full regression results on file with the author.

R^2:0.698
Adj. R^2:0.642
Std.Err.:1.635
$F_{(60,324)}$:12.46
p-val:0.000
N: 385

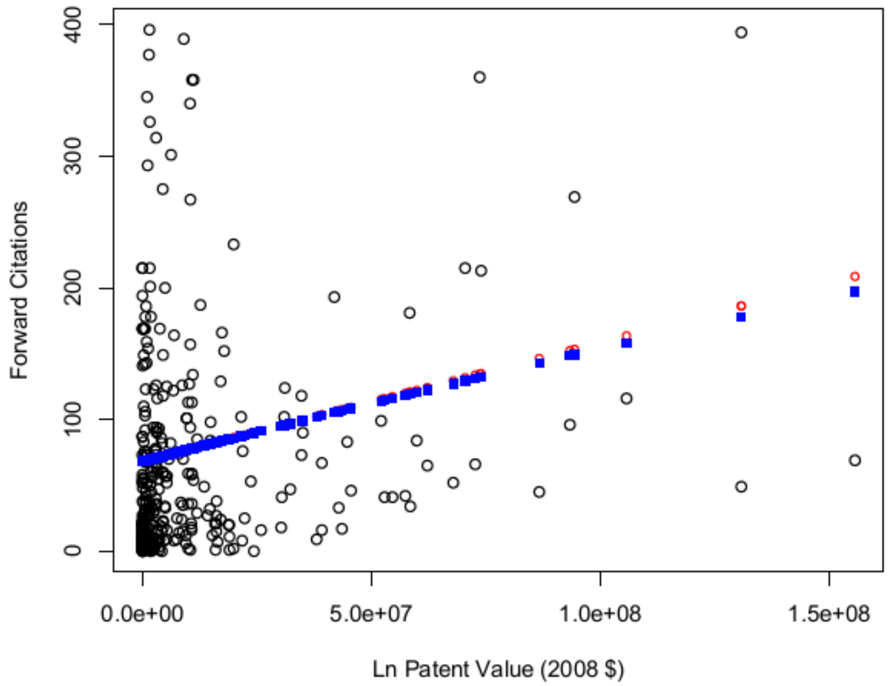
APPENDIX I – CITATION-TO-VALUE PLOTS:

Citations to Value (bucketed values)



(Avg. FC in green)
(linear trend in red)
(linear+quadratic in blue)

Citations to Value (log value scale)



(FC in black)
(linear trend in red)
(linear+quadratic in blue)

APPENDIX J – CITATION-TO-VALUE REGRESSION:

Variable	Coeff.	Std.Err.	t-val	Pr(>|t|)	Signif.
Patent Value	2.34e-8	4.42e-9	5.31	1.96e-7	***
Patent Value Squared	4.77e-17	1.29e-17	3.67	2.60e-4	***

Dependent variable is the log of Dec. 2014 citation counts.

R^2:0.105
Adj. R^2:0.0998
Std.Err.:1.492
$F_{(2,366)}$:21.4
p-val:0.000
N: 369