

# Texas Law Review

## *See Also*

### Response

#### Paving the Path to Accurately Predicting Legal Outcomes: A Comment on Professor Chien's *Predicting Patent Litigation*

Jay P. Kesan,<sup>\*</sup> David L. Schwartz<sup>\*\*</sup> & Ted Sichelman<sup>\*\*\*</sup>

Over 100 years ago, Oliver Wendell Holmes famously stated that “[t]he prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law.”<sup>1</sup> Unfortunately, neither Holmes nor his legal contemporaries possessed the tools to predict legal outcomes by using statistics.<sup>2</sup> Rather, they relied on ad hoc heuristics derived from case law, treatises, and professional knowledge to do so. Lawyers and legal scholars have largely continued to use these anecdotal approaches. Yet, with the advent of low-cost, high-speed, large-storage computing, Holmes’s vision of

---

\* Professor and H. Ross & Helen Workman Research Scholar, University of Illinois College of Law.

\*\* Assistant Professor of Law, Chicago–Kent College of Law.

\*\*\* Associate Professor of Law, University of San Diego School of Law.

1. Oliver Wendell Holmes, Jr., Supreme Judicial Council of Mass., *The Path of the Law*, Address Before the Boston University School of Law (Jan. 8, 1897), in 10 HARV. L. REV. 457, 461 (1897).

2. Although statistical methods—indeed, econometrics—were certainly available as a method to legal scholars, it appears that no significant efforts were made in the legal academia to use these approaches until the 1970s and 1980s. See MICHAEL O. FINKELSTEIN, *QUANTITATIVE METHODS IN LAW: STUDIES IN THE APPLICATION OF MATHEMATICAL PROBABILITY AND STATISTICS TO LEGAL PROBLEMS* (1978) (representing an effort to study case outcomes using statistical methods and to apply statistics in an objective effort to understand the law); Roderick A. MacDonald, *Access to Civil Justice*, in *THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH* 494–95 (Peter Cane & Herbert M. Kritzer eds., 2010) (presenting the history of the use of empirical methods in the legal academy). See generally MARY S. MORGAN, *THE HISTORY OF ECONOMETRIC IDEAS* (1990) (recounting the econometric approaches of the late nineteenth century and onward).

law-as-prediction-by-*lawyers* is being supplemented by law-as-prediction-by-*legal modelers*.

Unlike lawyers, legal modelers rely on quantifiable characteristics present in large datasets to make predictions about legal outcomes. For instance, for lawyers, the case citations present in opinions (e.g., “152 U.S. 688”) merely provide links to page numbers in case reporters that a lawyer may reference for a broader or deeper understanding. Yet, for legal modelers, each citation does not link simply to one page in another case, but to a web of opinions connected to each other through stacked sets of citations.<sup>3</sup> Case A cites Case B cites Case C, yet Case A does not cite Case C directly, and none cite Case D. By stringing together which cases directly cite each other, indirectly cite each other (and at what “degree of separation”), and do not cite each other, case citations do not merely provide references to background information. Rather, citations construct “legal networks” (much like “social networks”) that can be used to predict which cases a judge in any given case is likely to rely on, discredit, and wholly ignore, as well as whether certain judges are likely to agree or disagree with other judges.<sup>4</sup> This sort of quantified knowledge can be used to help predict legal outcomes—sometimes much better than the lawyers can predict. In a recently staged competition between legal academics and an expert algorithm in predicting Supreme Court outcomes, the algorithm won easily, predicting 75% correct outcomes, compared to 59% for the academics.<sup>5</sup>

In her well-written and highly original paper, *Predicting Patent Litigation*, Professor Colleen Chien takes the discipline of legal modeling one step further by taking it one step back (in time, that is). Specifically, Professor Chien departs from traditional models of predicting outcomes merely to ask whether we can predict the occurrence of patent litigation itself.<sup>6</sup>

Although scholars have examined whether certain “inherent” patent characteristics, like the number of claims,<sup>7</sup> number of citations,<sup>8</sup> and so forth, are correlated with the incidence of patent litigation, what makes Professor Chien’s article novel is that she examines a wide variety of after-acquired

---

3. See generally Thomas A. Smith, *The Web of Law*, 44 SAN DIEGO L. REV. 309 (2007) (discussing the networked nature of legal opinions).

4. See *id.* at 349–50 (positing that a legal network structure can be used to predict which cases will rise “above the rest”).

5. See Andrew D. Martin et al., *Competing Approaches to Predicting Supreme Court Decision Making*, 2 PERSP. ON POL. 761–64 (2004) (demonstrating the success of the author’s model at predicting case outcomes, and explaining the relative strengths and weaknesses of legal experts and legal models in predicting outcomes).

6. Colleen V. Chien, *Predicting Patent Litigation*, 90 TEXAS L. REV. 283, 287 (2011).

7. Jean O. Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: A Window on Competition*, 32 RAND J. ECON. 129, 140–41 (2001).

8. John R. Allison et al., *Valuable Patents*, 92 GEO. L.J. 435, 438 (2004).

characteristics of the patent to predict the onset of litigation.<sup>9</sup> These characteristics include changes in patent ownership,<sup>10</sup> investment in the patent,<sup>11</sup> reexamination,<sup>12</sup> securitization,<sup>13</sup> and forward citations.<sup>14</sup> Using multiple regression techniques, Professor Chien then derives correlations between these characteristics and whether patent litigation occurred.<sup>15</sup> She finds statistically significant correlations for all of the characteristics.<sup>16</sup> Importantly, Professor Chien then develops a model that can be used to predict the likelihood of whether a given patent will be litigated.<sup>17</sup> She asserts that these after-acquired characteristics, coupled with inherent characteristics, lead to a significantly superior model when compared to a model based entirely on inherent characteristics or none at all.<sup>18</sup> Unfortunately, Professor Chien does not compare her results to a sample of patent attorneys attempting to make the same large-scale predictions, but her model may very well outperform them. Whatever approach would be more accurate, Professor Chien's would clearly be cheaper and quicker.

As scholars in the field like Professor Chien—not to mention the scores of private companies using these techniques—apply these approaches to more and more questions in more and more disciplines, the nature of law as we know it will change quite dramatically. As such, Professor Chien's work is an important brick in a growing foundation of empirical legal studies (basically, the use of statistical methods to analyze the law)<sup>19</sup> and computational legal studies (basically, the use of computer methods and models to analyze the law).<sup>20</sup>

In the remainder of our Response, we first address what we perceive as several shortcomings in Professor Chien's analysis and then make recommendations for application of her approach and further study.

---

9. See Chien, *supra* note 6, at 287 (explaining that the author's study evaluates characteristics that a patent acquires over its lifetime, among other things).

10. *Id.* at 301–04.

11. *Id.* at 304–06.

12. *Id.* at 305–06.

13. *Id.* at 306–08.

14. *Id.* at 308.

15. See *id.* at 314–16 (explaining the author's methodology for determining correlations among the studied variables).

16. *Id.* at 317–18.

17. See generally *id.* at 320–26 (explaining the time-series model for predicting whether a particular patent will be litigated).

18. See *id.* at 322 (“The model that included both intrinsic and acquired characteristics was the most precise, generating only 406 false positives.”).

19. See generally Theodore Eisenberg, *Why Do Empirical Legal Scholarship?*, 41 SAN DIEGO L. REV. 1741 (2004) (recounting the rise and importance of empirical legal studies).

20. See Daniel Martin Katz et al., *Reproduction of Hierarchy? A Social Network Analysis of the American Law Professoriate*, 61 J. LEGAL EDUC. 76, 79 n.9 (2011) (“Computational legal studies is a sub-field dedicated to applying tools from computer science, applied physics, informatics, complex systems and applied mathematics to help enrich positive legal theory.”).

## I. A Critique of Professor Chien's Methodology and Policy Recommendations

As we mentioned earlier, we believe that Professor Chien has conducted an innovative study with important results. Nonetheless, there are limitations in the methodology that cast doubt on some of the policy recommendations that she makes.

### A. *Methodological Strengths and Gaps*

Professor Chien's general empirical approach of using a logistic regression is sound in major respects, and as we explained, her focus on after-acquired characteristics is not only original but a substantial contribution to the legal modeling literature. Nonetheless, we believe that her analysis could be strengthened in a number of ways.

1. *Data Gathering.*—There are a few shortcomings to Professor Chien's dataset that make her findings potentially less reliable as predictors of patent litigation than might appear at first pass. First, her dataset is limited to patents issued in 1990.<sup>21</sup> Although we recognize that Professor Chien wanted a set of patents that had fully expired, the characteristics of litigated patents today may vastly differ from those patents. The rise of nonpracticing entity (NPE) business models occurred primarily from the mid- to late 1990s<sup>22</sup> and has continued at an expanding pace since then.<sup>23</sup> A similar trend line can be found for operating companies expanding their patent portfolios for defensive reasons.<sup>24</sup> Thus, the characteristics of the 1990 patents associated with litigation—for example, the frequent assignment of litigated patents—may simply reflect the ascendancy of these business models in the 1990s and 2000s. If this is so, then litigated patents issued in the last five to ten years may be associated with fewer assignments, securitizations, and the like than those in Professor Chien's 1990 set. Related, there have been significant changes in the technology classes of patents asserted in litigation over the last twenty years. Because licensing and litigation tend to differ markedly across industries, Professor Chien's reliance on 1990 patents may fail to measure the changing characteristics associated with changes in the types of technologies asserted in patent actions. To be fair, Professor Chien

---

21. *Id.* at 309.

22. See Raymond DiPerna & Jack Hobaugh, *When to Litigate: Rise of the Trolls*, in LANNING G. BRYER ET AL., *INTELLECTUAL PROPERTY STRATEGIES FOR THE 21ST-CENTURY CORPORATION: A SHIFT IN STRATEGIC AND FINANCIAL MANAGEMENT* 143, 144 (2011) (describing the rise of the NPE model in the wake of the technology bubble of the mid- to late 1990s).

23. Colleen V. Chien, *Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents*, 87 N.C. L. REV. 1571, 1604 fig.2 (2009).

24. *Id.* at 1582–83.

recognizes these shortcomings in her article.<sup>25</sup> Yet, she draws the conclusion that because of them, “the results presented [in her article] may *understate* the relationships that currently exist between certain patent traits and litigation.”<sup>26</sup> Without the missing data, we do not see how Professor Chien can confidently rule out that her article *overstates* those relationships.<sup>27</sup> As such, we think Professor Chien’s analysis could have been strengthened by examining the characteristics of more recent patents, taking account for the unexpired patent term.

Second, even assuming that the patents used by Professor Chien were randomly selected from the relevant 1990 patents, we are not certain that Professor Chien’s sample is large enough to be representative of the entire population—at least when the control group is matched by patent class. Professor Chien’s sample size is only 2,636 patents, including 659 litigated patents and 1,997 control group patents.<sup>28</sup> The total number of in-force patents is presently around two million,<sup>29</sup> with approximately 30,000 of those likely to wind up in litigation.<sup>30</sup> Professor Chien’s 659 randomly selected patents of 30,000 in the total population is theoretically enough to ensure a sampling error of 5% at a 95% confidence level, if all patents were equally likely to be litigated.<sup>31</sup> Professor Chien recognizes that litigation rates may differ by technology and consequently and appropriately matches her

---

25. See Chien, *supra* note 6, at 309 (acknowledging that “more recent changes,” including inter partes reexamination and the “patent troll” phenomenon, were not captured by her model).

26. *Id.* at 310 (emphasis added).

27. Professor Chien also notes that she excluded patents held by Ronald A. Katz Technology Licensing L.P. from her sample. *Id.* at 309 n.173. We question whether the exclusion is appropriate. Professor Chien cites to the study of John Allison, Mark Lemley, and Josh Walker to support her exclusion. See *id.* (citing John R. Allison et al., *Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents*, 158 U. PA. L. REV. 1, 20 & n.39 (2009)). However, they merely excluded the Katz patents from the “small entity” count because those patents were not filed with small-entity status. Allison et al., *supra*, at 20. They did include all of the Katz patents in their statistical analysis. See *id.* at 26 (“We considered removing his suits from our results but decided against it, as the most-litigated patents are all outliers in some sense and removing Katz would skew our data. Katz is a product of the current patent system, and the Katzes of the world should be considered in evaluating the effects of that system.”). If the Katz patents are outliers and exclusion is appropriate, then the excluded data should be replaced—by mean substitution, regression substitution, or another approach—rather than excluded. See generally Xintao Wu & Daniel Barbará, *Modeling and Imputation of Large Incomplete Multidimensional Datasets*, 2454 LECTURE NOTES COMPUTER SCI. 286 (2002) (proposing a two-part model for predicting omitted values in datasets). Professor Chien should have at least reported the number of patents excluded and the effect the exclusion had on her results.

28. Chien, *supra* note 6, at 309.

29. For a collection of data confirming this result, see *Statistics on Patents*, WORLD INTELLECTUAL PROPERTY ORGANIZATION, <http://www.wipo.int/ipstats/en/statistics/patents/> (last updated Jan. 2011).

30. See Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSP., Spring 2005, at 75, 79 (“Only 1.5 percent of all patents are ever litigated, and only 0.1 percent are litigated to trial . . .”).

31. To confirm this calculation, see *Calculators*, SMARQUEST, <http://www.smarquest.ro/en/resources.html>.

litigated patents with control patents by USPTO technology class.<sup>32</sup> Yet, Professor Chien matches each of the 659 litigated patents to only three control patents issued in 1990 in the same primary patent class.<sup>33</sup> When a group is segmented into a variety of subgroups, a much larger random sample may be necessary to achieve representativeness. Here, if one particular patent class contains 1,500 issued patents in 1990<sup>34</sup>—and there is only one litigated patent in that class—then just choosing three of the 1,499 potential control patents would result in a *60% sampling error at even a 95% confidence level*.<sup>35</sup> Of course, the large patent classes may account for a larger share of litigation, which would increase the number of control patents in those classes, but surely some large classes (e.g., class 250, “radiant energy”) comprise few infringement actions. Thus, in order to properly ensure a representative control group for the litigated patents, it appears Professor Chien should have chosen many more matched patents for at least some patent classes with large numbers of issued patents per year.

2. *Empirical Methodology*.—Like our critique of Professor Chien’s dataset, we also point out several weaknesses in her methodology. One key, practical objective of work along the lines of Professor Chien’s is to develop the ability to predict and therefore, effectively purchase patent rights at an early stage when they may be available at a significant discount to preempt more costly access in the future. The efficiency of such an exercise depends on the extent to which a model is overinclusive or underinclusive in its prescriptions regarding which patents should be bought. In this regard, some of the variables that Professor Chien chose for her model may not be useful for predicting which patents should be licensed or purchased at an early stage. For instance, Professor Chien includes control variables for whether the patent was assigned and whether the patent was in reexamination.<sup>36</sup> We suspect that many times the patent is assigned only *after* the decision to litigate has been made. In these cases, the patentee is merely preparing for litigation by providing for an appropriate corporate entity to bring suit.<sup>37</sup> If

---

32. See Chien, *supra* note 6, at 309 (“For each litigated patent, I randomly selected an additional three patents issued in the same year and assigned to the same first-listed technology class, creating a matched-pair set.”).

33. *Id.*

34. Some subclasses include more than 1,500 patents in a given year. *Patenting by Geographic Region (State and Country), Breakout by Technology Class*, U.S. PAT. & TRADEMARK OFF., [http://www.uspto.gov/web/offices/ac/ido/oeip/taf/clsstc/allstcl\\_gd.htm](http://www.uspto.gov/web/offices/ac/ido/oeip/taf/clsstc/allstcl_gd.htm) (last modified Apr. 1, 2011).

35. To confirm this calculation, see *Calculators*, SMARQUEST, <http://www.smarquest.ro/en/resources.html>.

36. Chien, *supra* note 6, at 303, 306.

37. See Colleen V. Chien, *From Arms Race to Marketplace: The Complex Patent Ecosystem and Its Implications for the Patent System*, 62 HASTINGS L.J. 297, 319 (describing the phenomenon whereby entities assign patents to shell companies and sue companies that are utilizing competing inventions).

this is true, then an assignment generally speaking does not predict that the patent will be in litigation. Instead, the patent has already been selected for litigation, and attempts to purchase the patent at a discount will fail. Similarly, *ex parte* reexaminations may be filed immediately before litigation, when the parties are discussing potential licensing, for example.<sup>38</sup> Thus, by the time the reexamination is commenced, the patent already has been selected for litigation. In short, there is more of an endogeneity problem than Professor Chien acknowledges, as litigation may effectively predict these variables instead of these variables predicting litigation.<sup>39</sup>

We are also unclear on how Professor Chien coded and generated the data for her time-series model. She indicates that for the time-series model she used only “[c]haracteristics [a]cquired [p]rior to [l]itigation.”<sup>40</sup> This data is time-consuming to properly generate. For example, forward citation data should only include citations to litigated patents that occurred *before* litigation commenced.<sup>41</sup> The readily available forward citation information includes all forward citations during the life of the patent. Thus, for each litigated patent, the number of forward citations must be manually adjusted to exclude citations that occurred after the filing of litigation.<sup>42</sup> Professor Chien’s description of her methodology does not provide sufficient detail for us to confirm that she followed this protocol.

In addition, the kind of entity that owns the patents at issue may be a choice that reflects business- and tax-related decisions made by a party. The type of entity chosen may be informed by asset-transfer issues, financing decisions, beneficial state and federal tax treatment, and the like. And the decision making and control may exist not in the legal entity that owns the

---

38. *See* 35 U.S.C. § 302 (2006) (providing that a party may request *ex parte* reexamination of any patent claim “at any time”).

39. One of Professor Chien’s models includes characteristics acquired over the entire patent life. *See* Chien, *supra* note 6, at 329 app.a (listing Professor Chien’s variables methods with respect to intrinsic and acquired patent characteristics). We have concerns as the model includes characteristics, such as the filing of a reexamination request, which occur after the filing of litigation. As Professor Chien is relying upon the model to predict litigation, this regression model is methodologically unsound because it measures a “predictor” of patent litigation that occurred after the fact of litigation. In other words, in these cases, the filing of a reexamination is not likely to be a true independent variable that can be used to predict whether litigation is likely to occur for any given patent.

40. *See id.* (explaining this methodology).

41. This adjustment to correct forward citations should not be confused with the adjustment Professor Chien states that she made to exclude self-citations from forward citations. *See id.* at 308 & n.172 (“[T]he present analysis adjusts the number of forward citations by excluding citations that have at least one inventor in common with the cited patent.”).

42. Similarly, to ensure the reexaminations were filed before litigation, one would need to review the file wrapper from each of the reexamined patents to determine the date when the initial request for reexamination was filed. Chien notes that she utilized LexisNexis to ascertain whether a reexamination certificate had been issued. *Id.* at 312 & n.198. Because reexamination certificates only issue after the conclusion of a reexamination, her data necessarily omits any reexamination proceeding that was filed before, but concluded after, the onset of litigation.

patents, but instead in a parent or a related company. For example, a large manufacturing company may be the parent of an IP holding company. Yet, executives at the parent company often decide whether to litigate and, effectively, control all aspects of the litigation. Because ownership and decision making may rest with different entities, rigidly coding the patent owner as belonging to a particular type of entity may not capture the nuances of a given situation.

We have concerns that unobserved variables—specifically related to the type of patent owner—may compromise Professor Chien’s model. Because we believe that the unobserved variables are correlated with the observed variables, then the statistical inferences from the model may be questionable. Professor Chien herself repeatedly mentions the importance of the type of patent owner in predicting the possibility of litigation.<sup>43</sup> Yet, the variables that Professor Chien relies upon, such as assignments and small-entity status, are imperfect proxies and do not fully capture the type of patent owner. Professor Chien’s data lack other related and important “patent ecosystem” variables.<sup>44</sup> These unobserved patent ecosystem variables include, for example, whether the patent has been embodied in a product or service by the patent owner or another party, whether the patent is part of a portfolio of related patents owned by the patent owner, and the like. These variables directly affect the patent owner’s ability to trade effectively and perhaps balance the equities in a negotiation, dispute, or litigation in order to reach a pre-litigation settlement or a settlement soon after the initiation of a lawsuit. The ability to trade on multiple fronts may be unrelated or partially related to the type of patent owner. To the extent that these patent ecosystem variables that directly affect patent value and patent litigation are not taken into account, the statistical inferences that are drawn may be incomplete.

Finally, the article could have been improved by a few examples outlining the significance of the variables in the logistic regression. The information provided in the appendix includes the betas from the regressions.<sup>45</sup> Because the regressions are logistic regressions, the betas themselves are not meaningful to most readers unless converted into a different format. We would like to know what percentage change in an independent variable in the model—such as reexamination, maintenance fees, size of patent owner, etc.—resulted in what percentage change in the probability of the patent being litigated.

---

43. *See, e.g., id.* at 298, 319, 325 (reflecting Professor Chien’s argument that the identity of the patent owner is statistically relevant to whether the patent will be litigated).

44. *See* Chien, *supra* note 37, at 355 (generalizing observations regarding the environment within which patent litigation occurs).

45. Chien, *supra* note 6, at 329 app.a.

### B. Normative Inferences

Relying on her empirical results, Professor Chien makes several policy recommendations. Given the large number of variables involved in determining optimal policy prescriptions in patent law, we think it is generally quite difficult to do so from narrowly focused empirical studies. Here, we illustrate our beliefs by raising doubts about one of Professor Chien's normative conclusions: to require the real owners of patents to publicly disclose their identities. Specifically, Professor Chien contends that "if patents provide the right to exclude, the public is entitled to know who might do the excluding. Under the current system of recordation, accused infringers may have to wait until litigation to identify 'the real party in interest.'"<sup>46</sup> Her suggestion is based upon her finding that an understanding of patent owner characteristics aids in predicting whether a patent will be litigated.<sup>47</sup> Of course, predicting whether a patent is litigated is but one of many policy inputs into this difficult issue.

The more difficult it is to alienate patent rights, then the less value patents will deliver to inventors. Aside from desiring to mask its identity to make it more difficult for a potential infringer to predict litigation, a purchaser of a patent may wish to remain anonymous so as not to alert competitors to its long-term business plans or simply to lower the price it pays when purchasing a patent by keeping sellers in the dark about its overall assets. Moreover, in other areas of law, there is no similar transparency of ownership requirement. For instance, in real property, the use of shell holding companies to hold assets is widespread, and there is no mandatory "true owner" disclosure requirement.<sup>48</sup> Related, a party does not need to know the true owner of real property to avoid trespassing, which is roughly analogous to patent infringement. This would be true even if one could establish a statistically significant relationship between the true owner of real property and whether there will be a lawsuit alleging trespass. In this regard, a mandatory disclosure provision may encourage potential infringers to ignore patents. In particular, potential infringers may be more willing to disregard patent rights and infringe without fear, especially of patents owned by those empirically unlikely to litigate. We do not believe further incentives to disregard patent rights necessarily makes for good policy. Granted, there are "boundary" problems in patent law not present in real property law that might counsel more in favor of transparency in the patent law context. Yet, the relative benefits and costs of transparency are not entirely clear in our view, and Professor Chien's data does relatively little to solve the problem.

---

46. *Id.* at 327 (quoting FED. R. CIV. P. 17).

47. *Id.* at 326.

48. *See* HAROLD S. BLOOMENTHAL & SAMUEL WOLFF, 3A SECURITIES AND FEDERAL CORPORATE LAW § 6:69 (2d ed. 2009) (explaining the law surrounding shell companies and noting that they may hold property assets).

Thus, even if Professor Chien is correct in her empirical observations, we do not think her proposal naturally follows.

## II. Improving the Predictability of Patent Litigation

Professor Chien's article opens numerous avenues for future work. Her approach can be further refined and optimized in a number of ways to better model which variables predict patent litigation. Furthermore, the model can be extended to take into account more patent-related characteristics. Below, we briefly outline several avenues.

### A. *Refinements to Existing Data*

No doubt, there are many additional variables that could be included to structure a more robust model for predicting patent litigation, and we discuss some of those variables below. Here, we describe several incremental changes to the data that Professor Chien has already gathered, which we suspect may improve her model.

1. *Patent Ownership.*—Professor Chien divides the patent owner into two types: small and large entities. She also includes whether the patent was assigned (using PTO records)<sup>49</sup> and whether a patent holder ever changed from small-entity status to large-entity status (or vice versa).<sup>50</sup> While this is helpful, we believe that data about the patent owner should be more precisely defined. For example, party size can be broken down into various additional categories, including by market capitalization, by whether the company is publicly traded, and by whether the owner is foreign or domestic. Assignment data also can be better defined, including by determining whether the assignments were to an entity in the same industry.<sup>51</sup> By so

---

49. Professor Chien states that there is uncertainty over whether the number of unrecorded PTO assignment records is significant and that the current evidence is inconclusive. Chien, *supra* note 6, at 311. We believe that the number of unrecorded reassignments is likely significant. The only systematic data on the issue we know of indicates that a substantial percentage of reassignments of patents and applications go unrecorded, at least for startup companies. See Stuart J.H. Graham et al., *High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey*, 24 BERKELEY TECH. L.J. 1255, 1275 (2009) (“[T]he USPTO records on patents reassigned to different entities after grant are notoriously incomplete.”). Moreover, anecdotal evidence indicates the same across a variety of patentees. See Robin Feldman & Thomas Ewing, *The Giants Among Us 10–11* (Sept. 6, 2011) (unpublished manuscript), available at <http://ssrn.com/abstract=1923449> (noting discrepancies between public information and USPTO assignment records). Finally, even the “test analysis” of 100 patents that Professor Chien mentions in her article, thirty had likely unrecorded assignments. Chien, *supra* note 6, at 311. While she argues that 30% may overstate the number, *id.*, we feel it may in fact understate the number because no one knows how many of the remaining seventy patents in her sample had other, unknown unrecorded assignments.

50. Chien, *supra* note 6, at 312.

51. We also note that changes from small-entity status to large-entity status may be due to reasons beyond assignment to a larger company. For example, small-entity status is lost if the patent is licensed to a company with more than 500 employees. See 13 C.F.R. § 121.802 (2010) (providing that only concerns with up to 500 employees are eligible for reduced patent fees). To the

refining the data that Professor Chien has already assembled, she can more precisely capture this sort of information about the patent holder.

There are strong reasons to believe that the identity characteristics of the patent holder affect the chance of litigation.<sup>52</sup> Professors Ball and Kesan have previously reported data suggesting that small and large patentees enforce their patents at different rates.<sup>53</sup> Ball and Kesan find some evidence that small entities only enforce their “valuable” patents—as measured by intrinsic patent characteristics—while medium and large entities enforce their “average” patents.<sup>54</sup> For this reason, among others, we believe that Professor Chien’s ownership variables do not fully capture the complexity of patent litigation.

2. *Reexamination.*—Professor Chien includes a binary variable of whether the patent was involved in reexamination. She finds that reexamination is a significant positive predictor of litigation. Reexamination data can be further categorized. For example, reexamination can be initiated either by the patentee itself, by third party requestors, or by order of the Director.<sup>55</sup> Identifying the requesting party may be important and may be useful in the model.

3. *Timing of Transfers of Ownership.*—The timing of any transfer or assignment of a patent is important. Professor Chien finds that assignment of a patent predicts an increased likelihood of litigation.<sup>56</sup> In her model, she treats the assignment as a continuous variable, apparently counting the number of times the patent was assigned.<sup>57</sup> It would be helpful to identify the time period during which the patent was assigned. Some assignments may be related to the lawsuit, such as assignments done nearly contemporaneously with filing the complaint. In those cases, the patent may have been purchased for the purposes of litigation. Alternatively, as we

---

extent that many of the changes in small-entity status are due to licensing of the patent (rather than assignment or other reasons), then Professor Chien’s results, if correct, may show that licenses predict litigation.

52. Chien, *supra* note 23, at 1600 & tbl.3; Gwendolyn G. Ball & Jay P. Kesan, *Transaction Costs and Trolls: Strategic Behavior by Individual Inventors, Small Firms and Entrepreneurs in Patent Litigation* (Ill. Pub. Law & Legal Theory Papers Series, Research Paper Series No. 08-21; Ill. Law & Econ. Paper Series, Research Paper Series No. LE09-005; 2009), available at <http://papers.ssrn.com/sol3/?abstractid=1337166> (demonstrating empirically that the disposition of patent cases depends upon the nature of the parties).

53. Ball & Kesan, *supra* note 52, at 2.

54. *Id.*

55. Approximately one-third of the reexaminations are initiated by the patent holder, and two thirds are initiated by a third party. U.S. PATENT & TRADEMARK OFFICE, EX PARTE REEXAMINATION FILING DATA (2011), [http://www.uspto.gov/patents/EP\\_quarterly\\_report\\_June\\_2011.pdf](http://www.uspto.gov/patents/EP_quarterly_report_June_2011.pdf).

56. Chien, *supra* note 6, at 317.

57. *Id.* at 329 app.a.

discussed earlier, a patent holder may form a separate subsidiary to take advantage of corporate liability protections in the event that the lawsuit is unsuccessful. These contemporaneous assignment and lawsuit filings are of reduced value to the community, as the patent is already being asserted in litigation. On the other hand, if the assignment occurred years before litigation, it could be of tremendous value to know if assignments predict litigation. Thus, further research into the timing of transfer is paramount.

### *B. Extensions of the Model*

We see numerous directions in which Professor Chien's model can be extended. For space reasons, we mention just a few here. Before doing so, however, we flag the open question of whether Professor Chien's results are replicable. As we noted earlier, Professor Chien's sample size of less than three thousand patents is relatively small—particularly when subgrouped into classes—and her sample only includes patents issued in 1990.<sup>58</sup> As such, before Professor Chien's model is expanded to more variables, other researchers should test whether her results hold across larger datasets.

After replication, the most obvious extension of Professor Chien's work is to recognize that patent litigation comes in a variety of shapes and sizes. Ideally, we could determine the judgment or settlement amount of each lawsuit. Unfortunately, it is not an easy matter to determine the value of a lawsuit as most patent lawsuits terminate with a settlement.<sup>59</sup> Nearly all of these settlements are confidential, and they are unavailable for general study.<sup>60</sup> Consequently, we must look to other avenues to recognize value.

Another useful metric is litigation costs or attorneys' fees. Some patent infringement lawsuits settle relatively quickly and with little discovery and presumably, only modest legal fees. Other lawsuits pend for years—with extensive motion practice—and culminate with a trial or other adjudication on the merits.<sup>61</sup> Professor Chien uses a rather crude binary-output variable: whether or not there was litigation involving the patent.<sup>62</sup> We suggest the model be extended to include a more complex breakdown of litigation outcomes. The complexity can be described either by the duration the case pending, the number of papers filed with the trial court, or some combination

---

58. *See supra* text accompanying notes 21, 28.

59. Jay P. Kesan & Gwendolyn G. Ball, *How Are Patent Cases Resolved? An Empirical Examination of the Adjudication and Settlement of Patent Disputes*, 84 WASH. U. L. REV. 237, 273–74 & tbls.4, 5 & 6 (2006).

60. *See* David Luban, *Settlements and the Erosion of the Public Realm*, 83 GEO. L.J. 2619, 2650 (1995) (observing the prevalence of civil settlements that are subject to gag orders and confidentiality agreements).

61. Kesan & Ball, *supra* note 59, at 282.

62. Chien, *supra* note 6, at 314 & n.212.

of the two.<sup>63</sup> While these variables are admittedly imperfect—sometimes lawsuits pend for long stretches of time without any court action and sometimes there are few papers filed but the litigants spend tremendous amounts of money on discovery or other litigation activities—they are better than Professor Chien’s current variables.

Next, not all patents are allegedly infringed by only a single party. As Professor Chien herself has previously discussed, patentees in many instances assert infringement against numerous accused infringers, often in a single lawsuit.<sup>64</sup> Alternatively, a patent may be asserted against numerous parties in separate suits. Professor Chien’s current model treats all patent litigation as the same.<sup>65</sup> Differentiating patents and lawsuits by the number of alleged infringers and number of associated suits would likely be fruitful.

Finally, with the implementation of the America Invents Act of 2011 (AIA) already in process, there are a number of new avenues for examining patent characteristics. For example, the AIA introduces a post-grant review process, whereby issued patents may be challenged by third parties.<sup>66</sup> Identifying those patents in post-grant review, along with outcome results from the review, may prove helpful in predicting the propensity of a patent to be litigated.

### III. Conclusion

Professor Chien’s article is important in the evolution of empirical studies of the patent system. Indeed, predicting *ex ante* which patents might be the subject of future litigation is a grand question in patent empirical work that is worthy of our best efforts, and we commend Professor Chien for studying this challenging topic. The concepts that she introduced will have continued vitality in the literature as her model is improved and refined. Her approach can be adapted to other features of the patent system, and we expect a large body of research to follow in her wake.

---

63. Kesan and Ball use both of these metrics in their study of adjudication of patent infringement disputes. Kesan & Ball, *supra* note 59, at 285 tbl.12 (number of papers filed); *supra* text accompanying note 61 (case duration).

64. Chien, *supra* note 37, at 1572. The America Invents Act contains a provision on joinder that may alter the practice of naming multiple unrelated defendants in a single lawsuit. Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284, 332–33 (2011).

65. *See* Chien, *supra* note 6, at 313–14 (describing the empirical study of “litigated patents”).

66. § 8, 125 Stat. at 316.