Texas Law Review
See Also

Response

On Predicting Patent Litigation

Lee Petherbridge, Ph.D.*

I. Introduction

Predicting Patent Litigation represents an important addition to an extensive body of work that relates patent characteristics to patent litigation. Among the earlier studies, perhaps the most well-known to legal scholars is *Valuable Patents*.[1] That study identified a number of "key characteristics of litigated patents"—for example, small-entity status of the filer, number of claims, length of time in prosecution, number of patent-family members, number of forward citations, and patent age.[2] With some exceptions, many of these characteristics are fixed when a patent issues.[3] Professor Colleen Chien supplements this body of work by focusing a study on the characteristics that litigated patents acquire after they issue.[4]

Professor Chien thus gives us some new information to consider when we think about how it is that a patent comes to join the rarified group of litigated patents. But the main purpose of *Predicting Patent Litigation* is not to facilitate hypothesizing about what forces cause a patent to join the ranks

---

* Professor of Law and Richard A. Vachon, S.J. Fellow. My colleagues Carlos Berdejo and Michael Guttentag deserve many thanks for helping me work through some of the issues that arise in trying to predict patent litigation. They are not responsible, however, for any errors I may have made.

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

2. *Id.* at 438.

3. *See id.* at 461 ("[O]f the measures we have found predictive of patent litigation, only two—age of the patent when the lawsuit is filed and citations received—are unknown to the applicant by the time the patent issues.").

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

of the litigated, or to facilitate hypothesizing about what happens to patents that are on a litigation trajectory. Instead, the article emphasizes the use of acquired patent characteristics to facilitate patent clearance: the process of resolving whether an innovation will be economically affected by patent rights owned by another.[5] In empirical terms, the article does not merely claim to offer a model that explains or describes a few new characteristics of litigated patents. Instead it aspires to provide a model that is practically useful for predicting whether in the future a patent is likely to be asserted against an innovation. This Response focuses on that claim because it provides a useful avenue into a discussion about predicting patent litigation generally and about the use of patent-document and patent-file data for patent-clearance purposes.

II. Predicting Patent Litigation

The article aims to enable a test that can usefully allocate a patent to one of two groups: patents that will in the future be litigated (and thus made subject to additional efforts to clear) or patents that will in the future not be litigated (and thus cleared because they may be ignored).[6] This sounds great, but just the idea of predicting which patents will be litigated and which patents will not raises some not-insignificant issues. One stems from the fact that patent litigation is a relatively rare event. An application of Bayes' Theorem helps to illustrate the problem.

Conventional estimates are that about 1% of patents are litigated.[7] Employing some generous assumptions[8] about any test's ability to predict whether a patent will be litigated, one can begin to discern some general problems that affect the task.

Let:

$$Pr(T+|L) = .5$$

---

5. *See id.* at 322 ("These results suggest that litigation-bound patents can be identified ahead of time. The differences between litigated and unlitigated patents, therefore, would seem to have not only descriptive but also predictive power. . . . This suggests that a company seeking to clear its rights could focus on fewer high-risk patents identified by the model and ignore others.").

6. *Id.*

7. Adam B. Jaffe, *The U.S. Patent System in Transition: Policy Innovation and the Innovation Process*, 29 RES. POL'Y 531, 548 (2000).

8. I assume, rather than use probabilities taken from the article, for two reasons. The first is that this Response is making a general point about the difficulty of predicting a rare event like patent litigation, so the analysis in this Part is not specific to the article's model. The second is that it is not clear for the article's best model what the probability is that a patent tests positive given that it will be litigated. Having assumed the chance of a true positive, I did not want to confuse the reader into believing that the numbers I selected were taken from the article, so I went ahead and assumed the chance of a false positive result as well. The assumed probabilities are intended to be very generous to the article. If the reader thinks that they are not sufficiently so, it should be easy enough for the reader to insert whatever probabilities he or she thinks are more appropriate.

Where Pr(T+|L) is the chance of a positive test (T+) given that the patent tested really is a patent that will be litigated (L). In other words, the assumption is that there is a 50% chance that a positive test result is a true positive.

Let:

$$Pr(L) = .01$$

Where Pr(L) is the chance that any patent randomly selected for testing is a patent that will be litigated (L). This number comes from the conventional estimate that around 1% of patents are litigated.

Let:

$$Pr(UL) = .99$$

Where Pr(UL) is the chance that any patent randomly selected for testing is a patent that will not be litigated (UL). This number is obtained by subtracting Pr(L) from 1.

Let:

$$Pr(T+|UL) = .1$$

Where Pr(T+|UL) is the chance of a positive test (T+) given that the patent tested is a patent that will not be litigated (UL). In other words, the assumption is that 10% of the positive results will be false positives.[9]

Let us assume next that there are about 2,500,000 patents that an innovator could test.[10] Put in a table, we can get a sense of the relative probabilities across a population of patents. $Test_{L(+)}$ and $Test_{UL(+)}$ represent the test outcomes using our assumed probabilities, while $True_L$ and $True_{UL}$ represents whether patents are actually litigated.

---

9. At one point *Predicting Patent Litigation* indicates that the chance of a false-positive result in its best model might be .45. Chien, *supra* note 4, at 321 fig.4. So the probability I have assumed for a false-positive result (.1) is likely to be quite generous. As noted earlier, however, I have not included probabilities estimated from the article in the above-the-line analysis because it is hard to figure out what they are, and to avoid confusion. *See supra* note 8. In case .45 is an accurate estimate, I do track its implications below the line throughout the discussion.

10. As the model in *Predicting Patent Litigation* predicts on the binary category of maintenance fee payments, it does not appear exclude patents that are no longer in force for failure to pay maintenance fees. *See* Chien, *supra* note 4, at 329 app.a (identifying maintenance fees for patents in force prior to first litigation as a binary variable). In 2009, there were nearly two million patents in force in the United States. World Intellectual Prop. Org., *Statistical Country Profiles: United States of America*, WIPO RESOURCES, http://www.wipo.int/ipstats/en/statistics/country_profile/countries/us.html (last updated Sept. 2011). Assuming the accuracy of this statistic, and applying conventional estimates for the number of patents that could be in force (viz. they are less than twenty years from filing), but for failure to pay maintenance fees, the universe of patents that could be tested is probably on the order of 3,500,000. Even with all of that explanation, however, the actual number of patents available for testing is a fairly small point. The problem is in the probabilities.

Table 1. The Impact of a Test Predicting Patent Litigation on a Population of Patents

$Test_{L(+)}$	$Test_{UL(+)}$	*Population of Patents*
12,500 *Pr(T+|L) x 25,000*	12,500	25,000 $True_{L= Pr(L) = .01}$
247,500 *Pr(T+|UL) x 2,475,000*	2,227,500	2,475,000 $True_{UL=Pr(UL)=.99}$
260,000	2,240,000	2,500,000

Recall that the goal of the test is to allocate patents to one of two groups. One is patents that will be litigated (and thus subjected to additional investigation to clear). The other is patents that will not be litigated (and are thus cleared). If the test comes back positive, what is the probability that it has returned a patent that will be litigated? In other words, what is Pr(L|T+)?

Pr(L|T+) = the number of true positives/the total number positive results returned by the test.

Thus,

$$Pr(L|T+) = 12,500/260,000 = .048$$

To put this in perspective, recall that the probability that any randomly selected patent is litigated is .01. If the preceding calculation is correct, the test—on the assumptions used—improves the probability of accurately predicting that a patent will be litigated to .048. The improvement in probability is the difference, or .038.

Put slightly differently, before the test is applied, there is a 1% chance that a patent randomly selected will be litigated. After applying the test to select a patent, the chance that a selected patent will be litigated increases to 4.8%. It is not clear to me that this level of improvement is effective enough to practically identify for companies those patents that they should focus more resources on or those patents that they are free to ignore. Perhaps it is, but it raises some concerns that are highlighted by Table 1.

The first is that the test produces a large number of false positives. On the assumptions laid out above, the test will correctly allocate 12,500 patents to the litigated group. Another 247,500 patents, however, will be incorrectly

allocated to the litigated group.[11] Thus, for each true positive identified by the test, there will be approximately twenty-one false positives (i.e., patents that test positive but will not be litigated). Unable to distinguish the true positives from the false positives, innovators will face the prospect of committing substantial resources to the clearance of patents that do not need to be cleared.

As long as the probabilities do not change, the false-positive problem is not likely to go away. Thus, even if we assume, arguendo, that the population of patents that needs to be tested can somehow be reliably (viz. without excluding relevant patents) reduced to well below 2,500,000,[12] false positives remain a legitimate problem. What is needed is a more specific and more sensitive test. The problem with developing such a test is that patent litigation is a rare event and, as discussed in the next Part of this Response, a complex event that likely involves factors beyond those apparent in a patent file.

The second concern is that the number of false positives is not the only problem. One could perhaps speculate that a patent clearer might consider taking on the cost of clearing large numbers of patents that it did not need to clear if in doing so it could be confident that it had cleared all of the patents that it did need to clear. Unfortunately, even a test as sensitive and specific as the one assumed above will incorrectly allocate 12,500 patents to an unlitigated (and test-cleared) group containing another 2,227,500 correctly allocated patents.[13] So even if a patent clearer takes on the cost of clearing all of the patents about which the test suggests it has reason to worry, half of the patents that it should have reason to worry about will be missed because they are buried in a large set of overwhelmingly unlitigated patents.

Predicting Patent Litigation, unfortunately, misses the chance to directly confront these real-world patent-clearance concerns,[14] which makes

---

11. If the false positive rate is closer to .45, as *Predicting Patent Litigation* seems to suggest that it might be, then the number of false positives—patents incorrectly allocated to the litigated group—would be on the order of 1,113,750. Pr(L|T+) would equal .011, the improvement in probability over randomness would be .001, and an innovator would have to clear about ninety false positives for every true positive identified by the test. While this is an improvement—recall that without a test, an innovator would have to clear ninety-nine patents for each true positive—it might not be enough of an improvement to make the model, or anything similar, practically useful. *Cf.* Chien, *supra* note 4, at 322–23 (explaining how the acquired factors are valuable for predicting things like litigation risk and portfolio value).

12. Because the problem stems from the fact that patent litigation is a rare and complex event, it does not matter much, for example, if the population of patents involved is 250,000 rather than 2,500,000. Any test to predict patent litigation will need to meet fairly exacting standards of specificity and sensitivity. Otherwise, the concerns expressed here are unlikely to dissipate.

13. If the false positive rate is .45, 1,361,250 patents will be correctly allocated. The explanation for the difference over Table 1 is that the unlitigated patents will have been incorrectly allocated to the litigated group.

14. Although it does provide a modest discussion of how well its specification predicts in its sample. *See, e.g.*, Chien, *supra* note 4, at 321 fig.4.

the article unconvincing on the point that there is much practical utility in trying to predict patent litigation using the variables that the article introduces to the field, or using patent document or file characteristics more generally. It is not, for example, apparent from the article that the model specified represents an improvement over having an industry-experienced in-house patent agent or attorney devote a small fraction of working time to occasional keyword and technology-class searches of patent office data. In addition, the lack of an analysis concerning the relationship between the article's statistical arguments and the litigation-risk and patent-clearance contexts makes much of the article's discussion of these issues seem misplaced. On these points, the article ultimately reads more like a wish or an assumption that large-patent-number data-driven patent clearance is practicable, than it does a written description of how it might be practiced.

On the other hand, if one is willing not to completely take the article on its own terms, and is thus willing to set aside somewhat the article's emphasis on using acquired characteristics (or other patent-document and patent-file data) to resolve whether at some point in the future an innovation is likely to be affected by a patent, the results presented do offer some useful insights. In particular, the article introduces evidence for what are, to my knowledge, some new factors that associate with litigated patents. These factors are intuitive and should be helpful to future work concerning, e.g., the value of patents, patent litigation, and patent clearance. In addition, assuming the results can be replicated, and assuming other advances in this area, some of which are implicated by the subsequent Parts of this Response, it is possible that some acquired variables discussed in the article might someday find a home in a model that might be helpful to innovators in sussing out economically relevant patents.

III. The Problem of Using Acquired Characteristics to Predict Patent Litigation

In this Part, I leave aside the concerns of the previous Part over the sensitivity and specificity of a test for identifying ex ante patents that will be litigated, although such concerns affect and are affected by the concerns addressed here. The issue developed here is this: How informative can the specified acquired variables be when it comes to predicting whether or not a patent will be litigated? *Predicting Patent Litigation* acknowledges this issue[15] but gives the topic somewhat shorter shrift than it deserves. The issue can be expressed in at least two ways. First, in a nontrivial number of cases,

15. *See id.* at 320, 321 (acknowledging that endogeneity effects "can prevent factors that are correlated with an outcome from having any predictive value," but reporting that application of a time-series model to the data to control for these effects revealed that "differences between litigated and unlitigated patents were observable prior to the litigated patent's first litigation" in every dimension considered).

and for a nontrivial number of acquired characteristics specified in the model, the direction of the predictive relationship between the variables specified and whether a patent will be litigated is "litigation then specified variable," rather than "specified variable then litigation." Alternatively, the concern might be expressed as one of a lurking variable: the idea that there is some variable that causes both the specified variable and litigation. An example will help to illustrate the problem.

Counsel at a major technology company has been asked to determine whether an innovation in which the company is considering an investment of millions of dollars is likely to be dominated by the patents of others if the company pursues the innovation. Using a model like the one presented in *Predicting Patent Litigation*, counsel screens patents and offers the opinion that there is little to worry about.[16] How good is the opinion?

To facilitate analysis assume, as *Predicting Patent Litigation* implies, that a determination that a patent will not be litigated based on the model of a sample not too different from the one used in the article is sufficient for counsel in the above example to offer an opinion that there is little to worry about. For simplicity's sake, assume also that an opinion might turn on the marginal predictive power of reexamination (the categorical presence or absence of a reexamination certificate in the patent file), although it seems clear that other specified acquired variables raise the same concerns.

Consider what might happen next. In view of the advice, the company decides to go ahead with the innovation, but it happens that there are patents that were cleared—again, to keep it simple, let us assume primarily because of the marginal predictive power of the absence of a reexamination certificate—that are plausibly infringed by the innovation. If at some point—perhaps when the innovation hits the market and is being madly consumed by the public—the details of the innovation become sufficiently publicly known that the patent owner has the option of pursuing a suit for infringement, was the initial determination that the patents were cleared a good one? Was the company free to infringe? Intuition suggests that the answer in many cases will be "no." The technology company's behavior will have changed the patentee's benefit and cost expectations for a patent-infringement suit; where a suit made no economic sense before, it makes good economic sense after the competitor starts marketing its successful innovation.

To finish this discussion, it must be obvious that reexamination may still happen in this hypothetical situation—unfortunately, only long after the risk-creating patents have been cleared, investments have been made, and the company is facing substantial rent diversion and possibly expensive patent

16. *See, e.g.*, *id.* at 322 (positing that a litigation-prediction model "could be useful to a company deciding whether to pursue research in [a particular technological field] or another field").

litigation. Why? The answer stems from the fact that decisions to litigate might often cause reexamination.

In one situation, reexamination happens because patent owners have identified an innovation against which they desire to assert patents. To scrub the relevant patents before listing them in a complaint, they put them into reexamination. When this happens, there are at least two reasons why competitors will have trouble predicting *ex ante* which patents they need to worry about. The first reason is that in some instances, it is already too late. The litigation has "left the barn," so to speak, and investments made in an innovation are irretrievably at risk. A second reason is that when reexamination is undertaken to strengthen patents for a complaint, there might be a very short period of time between the issuance of a reexamination certificate and the filing of a complaint. Accordingly, the observation of the patent file might have to happen within a very narrow window of time for the reexamination characteristic to assist in finding some troublesome patents.

Moreover, it is not just patentees that can put patents into reexamination. Competitors who anticipate being sued for patent infringement sometimes put patents that they anticipate will be asserted against them into reexamination.[17] The purpose in doing so can be to get the patent office to invalidate claims or force the patentee to narrow the scope of claims, hopefully to the point where the patent is less likely to be asserted against the competitor's innovation. For a competitor so situated, reexamination may not be that helpful in identifying patents that the competitor needs to worry about because it is the competitor's perception of litigation risk that is a substantial cause of reexamination.

It must be said that while reexamination might not be much help in situations like the ones just described, reexamination might be more useful as a predictor for other competitors (viz. those that are not the target of the first litigation of a patent). On the other hand, if most patents that are litigated are litigated only once, then it is possible that reexamination might not be a very useful predictor in a nontrivial number of cases.

The preceding discussion suggests that using reexamination and other specified acquired characteristics to predict patent litigation may be of limited real-world usefulness.[18] One way to articulate the concern is, as just discussed, to think of litigation causing reexamination, not the other way around. Another way of articulating the concern—and perhaps one that is more fruitful—is to imagine that there is a lurking variable, some kind of third variable that is causal to both reexamination and litigation and that is

17. For a discussion of the use of this process and potential abuses, see Raymond A. Mercado, *The Use and Abuse of Patent Reexamination: Sham Petitioning Before the USPTO*, 12 COLUM. SCI. & TECH. L. REV. 92 (2011).

18. Neither of these examples is meant to suggest that reexamination *cannot* be a part of a specification useful for identifying patents that will later be litigated.

confounding their relationship. I am not sure what it is, but it could be something like: a competitor decides to innovate in market space that a patentee believes could provide a valuable revenue stream. That sort of confounding variable could stimulate the patentee and competitor behavior (viz. putting patents into reexamination) described above. A more general implication of such a confounding variable, should it exist, is that patent litigation might significantly depend on substantial actions on the part of infringement defendants. If that is true, the practical utility of using patent-document and patent-file information to predict patent litigation might be limited, and that the problems laid out in the previous Part—concerning sensitivity and specificity of prediction—could be very difficult to overcome.

In this vein, I think that *Predicting Patent Litigation* could have benefited from an articulation of its theory of patent litigation. By considering a theory of patent litigation, the article might have developed a number of factors that lead to patent litigation beyond those that it collected. By laying bare some of those factors, even generally, the article could have given the reader a better sense of how much to expect from the acquired characteristics it uses.[19] In addition, by considering theory, the article might have revealed other variables that could have been not only specified but perhaps also modeled on the sample collected.[20]

Similarly, *Predicting Patent Litigation* could have benefited from a theory of patent clearance. The article takes the view that identifying patents that will be litigated allows patents to be excluded from further attention—and thus cleared—and also facilitates clearance of patents that might affect an innovation because it will inform innovators about whether they should risk infringement of a patent, license a patent, design around it, or abandon an innovation altogether. But even if one assumes away the problems discussed above concerning the difficulties involved in reliably identifying patents that will in the future be litigated, it is still unclear that such identification leads comfortably to clearance. For as the article recognizes, at least some scholarship has laid the problem of clearance not on the doorstep of whether the patent was collateralized or how many logged forward cites it has, but rather on the doorstep of poor patent notice: the idea that it is difficult to tell what scope a patent has and thus to judge the likelihood that a patent is not invalid or that an innovation might infringe vel non the patent.[21]

---

19. For example, theory might suggest that patent litigation is at least in part driven by factors that are external to a patent's characteristics (e.g., economic considerations or industry).

20. *Accord, e.g.*, Chien, *supra* note 4, at 321 & fig.4 (arguing that the acquired characteristics of litigated patents vary based on the identity of the patent holder).

21. *Id.* at 289–90 (discussing the scholarship of Professors Bessen and Meurer, who "contend that the patent system fails to provide clear notice, including in the context of patent clearance").

IV. Relating the Model to Patent Clearance

To elevate the claim in the article from, "Here are some variables that describe attributes of some patents that were litigated," which represents an interesting and valuable (albeit more modest) contribution in its own right, to the much more heroic "Here is a model that shows innovators what they need to do to determine what patents they need to worry about," I think the article has some obligation to show how the model employed relates to that real-world goal. Thus, even leaving aside the concerns raised in the previous Parts that have some bearing on this point, I think it is important that the article at least considers how the model it specifies, or something enabled by what it specifies, can be applied (even at a crude level) to some (even hypothetical) clearance situation. In the brief time I have had with the article, I have found this task somewhat challenging. This Part offers some examples of areas that might be given more attention in future work.

How should a patent clearer use the odds ratios (or coefficients)[22] determined by this model, or one like it? For example, the payment of maintenance fees is required to keep a patent in force,[23] and it is essentially only in-force patents that can be litigated in an action for infringement.[24] The payment of maintenance fees, therefore, seems like a theoretically terrific variable for inclusion in a specification designed to predict patents that will be litigated. When one looks at the Lifetime Model (including both intrinsic and acquired characteristics), maintenance fees (2.162) predict that the odds in the sample that a patent is litigated are 216.2% higher when a maintenance fee is paid up-to-date than when a maintenance fee is not paid up-to-date.[25] This, as the article points out, is a fine thing to know, but waiting for the end of a patent's life to measure maintenance payments is pretty useless if one is trying to make a prediction about the litigation risk presented by the patent during its life. To remedy this, the article also provides a time-series model,

---

22. *Predicting Patent Litigation* reports odds ratios rather than logistic coefficients, so I talk about odds ratios in the text. But a similar point can be made about logit coefficients.

23. 35 U.S.C. § 41(b) (2006).

24. *See id.* § 271(a) ("[W]hoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention *during the term of the patent therefor*, infringes the patent." (emphasis added)). There is a short grace period during which an owner may restore a patent that has expired for failure to pay maintenance fees. *See id.* § 41(b)(2) (creating a grace period of six months upon payment of an applicable maintenance fee and discretionary surcharge). In addition, if circumstances are exceptional, even after the grace period has expired, some patents may be restored. *Id.* § 41(c)(1). *But see* § 41(c)(2) (providing that the revival of a patent after the expiration of the grace period shall not affect the rights of parties who took otherwise-infringing action between expiration of the grace period and revival of the patent term).

25. Chien, *supra* note 4, at 329 app.a.

which appears to cut off the data for all of the patents in a matched set[26] at the time the litigated patent in a set is first litigated. The effect of this on the maintenance fees' odds ratio is, apparently, to turn it into a very large number—approximately 7×10^8, or 700,000,000. This suggests that in that model, maintenance fees predict a large increase in the likelihood of litigation (and produces a seemingly large standard error).[27] How, even hypothetically, should an innovator use the information from either of these models to make a decision?[28]

Another issue to consider in connection with how the specification applies to the real-world task of patent clearance implicates the sample used in the study. The specification is modeled with a sample of patents issued in 1990. What if there is something different about patents that will be litigated but that were issued in 2000, 2001, 2005, or 2007? Is the specification to use the same for these patents? How should a patent clearer model patents issued, for example, in 2009? The lifetime model will be unavailable, and the time-series model might mostly reflect the influence of patents that are litigated very close to their issue date. If those patents are special—perhaps because the patentee had them issued for the purpose of litigating—how should a patent clearer handle the rest of the cohort? If the sample of patents used here is not the ideal sample for patent-clearance purposes, what, in even broad strokes, would a better sample look like?

Does the specification presented work for patents directed to any industry (e.g., pharmaceutical, aerospace, information technology)? Does it work no matter the patent-owner-identity profile (e.g., nonpracticing-entity patent asserter, practicing-entity patent asserter, individual inventor)?[29] The matching technique employed matches on technology class, which might have some relationship to some of these variables, but intuition suggests

---

26. The sample date is defined as a random set of patents issued in 1990 and litigated. *Id.* at 309. These were matched with other patents, also issued in 1990, that had the same first-listed technology class. *Id.* Each litigated patent was matched with three unlitigated patents. *Id.*

27. *Id.* at 329 app.a (reporting the results of a time-series analysis conducted on a model that contained both intrinsic and acquired variables).

28. One possible future approach might be to decide that if a patent has expired for failure to pay maintenance fees, it cannot be litigated. For clearance purposes, this could get rid of a substantial number of patents, since empirical research has shown that approximately 16% are not renewed by four years after issuance, 37% by eight years, and slightly over 53% by twelve years. Kimberly A. Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521, 1531 tbl.1 (2005) (reporting the results of a study that looked at maintenance fee payments for all utility patents issued in 1991); *see also* Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341, 362 (2010) (reporting that patentees fail to pay fees on more than 60% of patents issued within twelve years). Then, a similar specification, although lacking the maintenance-fee variable, might still be used to evaluate acquired characteristics. Another approach could be to develop a specification for use with a sample that permits greater maintenance-fee variation on litigation. Matching on maintenance fees might be yet another option to consider.

29. *Cf.* Chien, *supra* note 4, at 325 fig.5 (arguing that the acquired characteristics of litigated patents vary based on the identity of the patent holder).

caution with respect to the extent to which a technology-class match can "fix" the effects of such variables. This concern is amplified by the form of regression model used, which seems to pool all of the matched sets, risking the possibility that much of what is being observed could be accounted for by things like industry-specific or owner-specific effects.

V. Conclusion

The patent system—and information about the patent system—is becoming more available and more digitalized. This trend rightfully encourages efforts to use information to make the patent system work better. *Predicting Patent Litigation* taps into this trend and makes an important contribution. It introduces what are, to my knowledge, some new acquired characteristics for the field to consider, including entity-size change, collateralization, and perhaps assignment[30] and transfer. It also gathers evidence that suggests the direction of the relationship between the characteristics and patent litigation in some circumstances. These are important findings that will facilitate hypothesizing about patents in a variety of contexts. And while the notion that the results presented in the article show the practicability of large-patent-number data-driven patent clearance, or show innovators how to figure out what patents they need to worry about, seems much too grand, Professor Chien has without a doubt contributed meaningfully to the study of patents, patent litigation, and the patent system.

30. In interesting contrast to the other characteristics studied, assignment predicts a fairly impressive decrease in litigation odds in the model. *Id.* at 329 app.a.