Jean O. Lanjouw and Mark Schankerman

Enforcing patent rights: an empirical study

Workshop paper

We thank the National Academy of Sciences and the Brookings Institution for financial support.

Enforcing Patent Rights: An Empirical Study

Jean O. Lanjouw and Mark Schankerman*

August 2002

* We thank the National Academy of Sciences and the Brookings Institution for financial support, and Derwent for generously providing access to their LitAlert database, which was critical to making this project feasible. We thank Bronwyn Hall, Adam Jaffe and Manuel Trajtenberg for their provision of some data inputs, Joe Cecil from the Federal Judicial Center and Jim Hirabayashi of the U.S. Patent and Trademark office for helpful discussions about the court and patent data, as well as Marty Adelman, Wes Cohen, Adam Jaffe and Kimberly Moore for useful comments. Maria Fitzpatrick provided excellent research assistance.

Enforcing Patent Rights: An Empirical Study

Abstract

We study the determinants of patent suits and settlements during 1978-1999 by linking information from the U.S. patent office, the federal courts and industry sources. Litigation risk is much higher for patents owned by individuals and smaller firms. Patentees with a large portfolio of patents to trade, or other characteristics that facilitate "cooperative" resolution of disputes, are less likely to litigate. But post-suit outcomes do not depend on these characteristics. These findings show that small patentees are at a disadvantage in enforcing their patent rights since their greater litigation risk is not offset by more rapid resolution of their suits. Our empirical estimates of the heterogeneity in litigation risk can help develop private patent litigation insurance to mitigate the adverse affects of high enforcement costs.

JEL Codes: K41, O34
Keywords: patents, litigation, settlement, R&D

Jean Olson Lanjouw
University of California, Berkeley and
The Brookings Institution
1775 Massachusetts Ave, NW
Washington DC 20036
Email: jlanjouw@brook.edu

Mark Schankerman
Department of Economics
London School of Economics
Houghton Street, London WC2A 2AE
United Kingdom
Email: m.schankerman@lse.ac.uk

1. Introduction

The central purpose of the patent system is to provide R&D incentives, but there is growing concern among scholars and the business community that fragmented patent rights and strategic patenting behavior are preventing firms from conducting R&D effectively. The perception is that litigation is becoming increasingly difficult to avoid, pushing up patent enforcement costs.[1] The fact that patent litigation grew extremely rapidly over the past two decades encourages this view. The number of patent suits rose by almost tenfold, with much of this increase occurring during the 1990's.

There is an extensive theoretical literature on why firms sue, but there is very little empirical evidence.[2] The literature on optimal patent policy has been studied without much empirical connection to the explosion of patent litigation over the past decade. In this paper we provide a solid empirical basis for evaluating what has happened in patent litigation and its implications for R&D incentives and patent policy. In particular, we show that the incidence of patent suits has *not* been rising, after controlling for the rapid increase in patenting itself and a shift toward more litigious technology areas, but also that it is concentrated in important ways in firms and patents with particular characteristics.[3]

The analysis provides empirical support for a key game-theoretic prediction in the context of the theory of litigation – that repeated interaction facilitates cooperation. Patent cases provide a unique opportunity to examine this prediction because case features are relatively well defined, allowing us to control for other aspects of disputes that might affect litigation risk, such as the economic value at stake. Since the basic factors affecting cooperation and settlement bargaining are general, our analysis of patent litigation can provide insights for other areas of study where it would be more difficult to disentangle them.[4]

---

[1] For example, see Heller and Eisenberg, 1998; Eisenberg, 1999; Hall and Zeidonis, 2001; Shapiro, 2001.

[2] Early theoretical work includes P'ng, 1983; Bebchuk, 1984; Priest and Klein, 1984; Spier, 1992.

[3] Since survey evidence indicates that the cost of patent suits that reach trial stage has grown substantially over the past decade (AIPLA, 2001), the overall burden of enforcement may well have risen.

[4] Cooter and Rubinfeld (1989) provide a good general discussion of the determinants of litigation.

In the patent context there are two main mechanisms by which disputants can settle without resort to the courts. The first is by "trading" intellectual property. This takes various forms, including cross-licensing agreements and patent exchanges, sometimes with balancing cash payments (Grindley and Teece, 1997). In fact, one motivation for accumulating patents may be to facilitate such trading (Hall and Zeidonis, 2001). The second mechanism that promotes settlement of disputes is the expectation of repeated interaction among patentees, either over time or in different markets. Game theory predicts that repeated interaction increases the incentive to settle disputes "cooperatively" – that is, without filing suits (e.g., Bernheim and Whinston, 1990; Tirole, 1994).[5]

Three of our key findings are consistent with the view that trading and repeated interaction are important for patent dispute resolution. First, having a larger portfolio of patents reduces the probability of filing a suit on any individual patent, conditional on its observed characteristics, and the quantitative effect is large. For a (small) domestic unlisted company with a small portfolio of 100 patents, the average probability of litigating a given patent is two percent. For a similar company but with a moderate portfolio of 500 patents, the figure drops to only 0.5 percent. This means that there are beneficial "enforcement spillovers" across patents within a given firm. Second, the portfolio effect is stronger for smaller companies, as measured by employment. For small firms, having a portfolio of patents is likely to be the key mechanism for avoiding litigation, whereas larger firms can also rely on repeated interaction in intellectual property and product markets to discipline behavior. Third, firms that operate in technology areas that are more concentrated (where patenting is dominated by fewer companies) are much less likely to be involved in patent infringement suits. We also find evidence of a threat value associated with having control over many patents in an area: firms having portfolios that are large *relative* to the disputants they are likely to encounter are significantly less likely to make use of the courts.

In theoretical models of litigation, settlement can occur at each stage of the legal process, beginning with the decision to file a suit and ending either with a post-filing settlement or

[5] There is very little econometric evidence to support this basic prediction. A notable exception is Siegelman and Waldfogel (1999) who construct measures of repeat play and find evidence that

adjudication at trial. We show that the *threat* of court action (suits) is the primary mechanism through which such 'sorting' among disputes occurs. Post-suit outcomes – the probability of settlement and the patentee win rates at trial – are almost completely independent of these characteristics, and this helps to mitigate the private (and social) costs of enforcement. Also lowering the cost of enforcement is the fact that post-filing settlement rates are high (about 95 percent), and that most settlement occurs soon after the suit is filed, often before the pre-trial hearing is held. These findings are good news: most cases are settled quickly, before extensive legal proceedings consume resources. The bad news is that the more frequent involvement of smaller patentees in court actions is not offset by a more rapid resolution of their suits.

Our findings imply that the enforcement process undermines the R&D incentives of small firms. They are the first direct evidence based on comprehensive litigation data, but they are consistent with earlier empirical studies. For example, Lerner (1995) shows that small firms avoid R&D areas where the threat of litigation from larger firms is high. Lanjouw and Lerner (2002) argue that the use of preliminary injunctions by large firms can discourage R&D by small firms, and this may apply to other legal mechanisms. Even if parties can settle their patent disputes without resorting to suits, the effective threat of litigation will influence settlement terms and thus, ultimately, the incentives to undertake R&D.

Both the advantages of patent portfolio and company size in settling disputes, and the heterogeneity in litigation risk, point to the potential value of patent litigation insurance. Insurance would enhance the ability of small firms and new entrants to enforce their intellectual property rights more effectively, and thus increase their bargaining power in licensing agreements and other forms of technology transfer. Currently there are a number of providers of litigation insurance in the United States and Europe. However, demand has been severely limited by high prices while, at the same time, profitability of insuring companies has been undermined by the widespread use of pooled prices. Concerns about the weak development of this market have recently been voiced in government and

---

reputation matters in various areas of litigation.

the public media.[6] Our empirical analysis could be used to develop insurance pricing schemes that recognize the heterogeneity of litigation risk.

The paper is organized as follows. Section 2 describes the construction of the data set and the empirical approach. Section 3 presents non-parametric evidence on how patent portfolio size and other firm and patent characteristics affect suit filings and their outcomes. Section 4 presents econometric analyses of the determinants of litigation for infringement suits and post-suit settlement, and illustrates the heterogeneity in the predicted probability of suits for different profiles of patent holders and patents. Section 5 briefly discusses implications for pricing patent insurance. Concluding remarks summarize areas for future research.

2. Data Description and Analytical Framework

We identify litigated patents from the LitAlert database produced by Derwent, a private vendor. This database is primarily constructed from information collected by the U.S. Patent and Trademark Office (PTO). The data used include 13,625 patent cases filed during the period 1978-1999. Each case filing identifies the main patent in dispute, although there may also be other patents listed. We use only the main patent in most of our analysis, which gives a total of 9,345 patents involved in our sample of suits. We then matched the Derwent data to information on all U.S. patent-related cases (those coded 830) from the court database organized by the Federal Judicial Center (FJC). This information includes the progress or resolution of suits as of the end of 1997.[7] Detailed information about the litigated patents and the universe of patents, and their owners, was drawn from a number of PTO databases.

When analyzing levels of litigation we adjust for both under-reporting and truncation in the Derwent data. The federal courts are required to notify the PTO of every case filing that involves a

[6] For an interesting policy discussion of past experience in this area, see the Danish Ministry of Trade and Industry (2001) report. A recent media story appeared in the U.K. Financial Times (September 14, 2001).

[7] The docket numbering was made consistent by hand, to allow merging. Discussions with the FJC indicated that some cases involving patents may be coded under other categories (e.g., the patent issue

U.S. patent, but under-reporting occurs in practice. Thus the patent office (and consequently Derwent) registers a subset of all patent cases. To estimate reporting rates, we take the number of cases filed according to Derwent divided by the number in the same year that are coded as a patent case by the FJC. Reporting rates stabilize in the 1990's at about 55 percent. We find no evidence of selection bias in the under reporting by the courts to the PTO: there are no significant differences between reported and unreported cases for a range of variables in the FJC database. Because we observe suit filings only through 1999, later cohorts of patents look like they are less litigated by construction. The lag structure for case filings for cohorts 1982-86 is used to adjust for this truncation.

We generated a "matched" set of patents to act as a control group. For each litigated patent, another was chosen at random from the set of all U.S. patents with the same application year and primary 3-digit U.S. Patent Classification (USPC) assignment. Therefore, the comparisons we present between litigated patents and matched patents largely control for technology and cohort effects.

In order to identify the type of patent suit, we manually matched the owner of each litigated patent to a party in the suit. A filed case is identified as an infringement suit if the patent owner is a plaintiff, and as a suit for declaratory judgement if the patent owner is a defendant (we do not analyze the latter in detail in this paper). This could be done for about 65 percent of the suits, of which about 85 percent are infringement suits. Cases where the patentee is not identified as one of the litigants are included as infringement suits in most of the analysis, since they are likely to be suits brought either by an exclusive licensee or by a subsidiary or head office of the patent-owning entity. (For further details regarding data construction, see Lanjouw and Schankerman, 2001b.)

We focus here on the effects of size and repeated interaction on the incidence and outcomes of patent suits, but we also control for other factors identified by Lanjouw and Schankerman (2001a). The latter include the number of potential disputes, the value of what is at stake, and the cost of prosecuting a suit in a U.S. court. Turning first to the control variables, we have the following characteristics for each litigated and matched patent:

---

may be part of a broader contractual dispute). This is also evident in the data where a small percentage of cases identified by Derwent are not in the FJC database.

Nationality of Patent Owner: We use the PTO designation of a patentee as domestic or foreign when it has been assigned to a company, and the address of the first listed inventor otherwise. Domestic patents account for 73.4 percent of the total. We expect domestic owners to have relatively lower costs of detecting and prosecuting infringements in the United States.

Technology Field: The USPC is a hierarchical, technology-based classification system. A patent examiner assigns to each patent one or more of the 421 USPC 3-digit classes. When a patent is assigned to more than one class, we use the first to identify the patent's field of technology. For most of the analysis we aggregate to eight broad technology groups: drugs, biotechnology, other health (mainly medical instruments), chemicals, electronics, computers, mechanical, and miscellaneous. The sum of a patent's unique 3-digit class assignments is used as an indicator of its technological breadth.

Number of Claims: The set of claims in a patent application delineates the boundaries of the property rights in the patent. The principal claims define the essential novel features of the invention, and the subordinate claims describe detailed features of the innovation claimed. The patentee has an incentive to claim as much as possible in the application, but the patent examiner can narrow claims before granting. We expect the number of claims to be associated with greater patent value, and thus more disputes.

Citations: An inventor must cite all related prior U.S. patents in his patent application, and a patent examiner is responsible for insuring that all appropriate patents are cited. Citations in the patent document help to define the property rights of the patentee. For each patent in the litigated and matched data, we obtained the number of prior patents cited in the application (*backward citations*) and their USPC sub-class assignments. We obtained the same information on all of the subsequent patents citing a given patent in their own applications, as of 1998 (*forward citations*). When a patent is assigned to a firm, we determine the share of forward and backward citations that are from the same firm (*self-citation*). Greater forward self-citation suggests that a patent protects an initial stage in a "cumulative innovation" developed by the patentee. Greater backward self-citation suggests that a patent protects a later stage in a cumulative chain. Both forward citation by others and self-citation suggest that a patent is more valuable and thus more likely to be involved in suits, while greater backward citation (including self-citation) indicates a lower-valued patent in a well developed area.

We now turn to our focus variables that capture the role of patent portfolio size, the size of potential disputants, and technology field concentration.

Ownership: We identify each patent owner as an individual, an unlisted company, or a listed company.[8] Individual and firm owners are indicated as such in the PTO data, and each firm is assigned a patent is given a company code by the PTO. To identify firms listed on the stock market, we use the concordance between PTO company codes and Standard and Poor's CUSIP identification codes developed by Bronwyn Hall and Adam Jaffe. This concordance is based on the 1989 industry structure. We call a patent-owning company "listed" if we can identify it as having a code at that time.[9] Unlisted companies are typically smaller than listed ones, but there is wide within-group variation. Individuals and listed companies are more predominantly domestic (81.0 and 95.6 percent, respectively) than unlisted companies (60.4 percent). We also break down listed firms into "large" firms (those with employment above the median of 5425) and "small" firms with employment below the median. Unless otherwise noted, we classify the nearly 40 percent of firms without employment data as large firms because they have similar litigation and settlement patterns.

Patent Portfolio Size: The PTO company codes allow us to construct a measure of the size of a patent owner's portfolio, as it looks during the period a patent would likely be involved in a dispute. The relevant portfolio is defined as the set of patents held by the patent owner having an application date within ten years in either direction of the patent in question. Notice that this portfolio size variable may differ across patents, for a given company. Not surprisingly, domestic listed companies tend to have larger portfolios – roughly a third of patents owned by domestic listed companies are in

[8] A small share of patents is assigned to institutions, such as universities, hospitals or governments. We treat these as unlisted companies.
[9] Two points are worth noting here. First, companies that merge after 1989 stop accumulating patent portfolios because their subsequent patenting is listed under a different (merged company) code. Second, any listed company that is started after 1989 will not have a CUSIP in our data and thus will be coded as an unlisted company.

portfolios in each of size groups 1-100, 100-900 and >900 patents. By contrast, about 90 percent of patents owned by domestic unlisted companies, and two-thirds of patents owned by foreign companies, are in portfolios with fewer than 100 patents.

Relative Size of Potential Disputants: We construct a measure of the asymmetry in portfolio size between a patentee and the "representative" disputant he can expect to face on each patent. Disputes will typically occur between firms engaged in similar lines of research. Such firms are also more likely to cite each other's patents as prior art. Therefore, we identify firms patenting in the same technology areas as a given patent's forward citations as the likely potential disputants for that patent.[10] We define size asymmetry as the firm's total portfolio size divided by a weighted average of the portfolio size of firms in classes from which its forward citations come. Formally, let Z_{cf} be the portfolio size for firm f in technology class c (including all patents since 1978) and $Z_{c\cdot} = \Sigma_f Z_{cf}$ all patents in the class. $Z_{\cdot f} = \Sigma_c Z_{cf}$ is the portfolio size for firm f, and $Z^*_{c\cdot} = Z_c / n_c$ is the average portfolio size of the n_c firms with patents in class c. The relative portfolio size of firm f for patent i is $R_{if} = Z_{\cdot f} / \Sigma_c w_{ci} Z^*_{c\cdot}$, where $w_{ci} = F_{ci} / F_{\cdot i}$ is the fraction of forward citations to patent i that falls into technology class c.[11] For a patentee suing for infringement, being relatively large confers greater threat power (e.g., holding cross-licensing of other patents hostage to this dispute) and this should facilitate settlement.

Technology Concentration: To measure of concentration in patenting activity in the technology area of each patent, we first calculate, for each 2-digit USPC class, a four-firm concentration index, measured as the patenting share of the top four firms. A firm's share is the size of its patent portfolio in that class divided by the sum of all firms' patents in that class. For each patent we then take a weighted average of the concentration indices for the different classes, where

[10] This identification is supported by an analysis of actual defendants. We compare the 3-digit classifications of all forward citations of the patent involved in a suit to the technology classes of patents owned by actual defendants in that suit. The share of classes that overlaps varies from 0.16 to 0.47, depending on the type of innovation. By contrast, the overlap for a random set of patents from the same cohorts is an order of magnitude smaller, ranging from 0.016 to 0.059.

[11] For a patent without any forward citations, the denominator of *Relsize* is set equal to the average portfolio size for other patents in the same 2-digit USPC class. For all individuals, and for about 900 cases where company patentees had only one patent, we set *Relsize* equal to zero.

the weights are defined as above. Formally, the index for class c is $C4_c = \Sigma_f Z_{cf} / Z_{c\cdot}$, where the sum is over the top four firms in terms of shares in that class. The weighted technology concentration index for patent i is then $C4_i = \Sigma_c w_{ci} C4_c$. If a company operates in more concentrated technological areas, it faces a greater chance of encountering other firms in patent disputes more than once. This expectation of repeated interaction should lower the litigation rate.

We analyze the litigation process in four stages: 1. the probability that a suit is filed; 2. the probability of settlement, conditional on a suit being filed (post-suit settlement); 3. the timing of any post-suit settlement; and 4. plaintiff (patentee) win rates, conditional on adjudication at trial.[12] If a patent dispute is settled before a suit is filed, we do not observe the dispute in the data. Thus a low filing rate can either reflect a low rate of infringement (disputes) or a high probability of pre-suit settlement. After a suit is filed, settlement can occur before the pre-trial hearing, after the hearing but before the trial begins, or during the trial. Otherwise, the trial concludes with a court judgement in favor of one of the parties.[13] We analyze the timing of case closure in terms of stages, rather than months, because patent lawyers have indicated to us that legal costs are more closely related to the stages a case reaches than to its actual length (which depends heavily on the availability of court resources and other external factors).

Information on trial win rates is relevant for assessing overall litigation risk, and thus pricing patent insurance. Such information is also useful for testing competing economic models of litigation (Waldfogel, 1998; Siegelman and Waldfogel, 1999). There are two main models: divergent expectations (DE) and asymmetric information (AI). In DE models (Priest and Klein, 1984), each party estimates the quality of his case with error (equivalently, the relevant court decision standard), and cases go to trial when one party is sufficiently more optimistic than the other. This occurs most

[12] Both parties can win. For example, a court may rule that an infringement occurred but strike down the validity of some of the patent claims. When a win for both is recorded, we count it twice, for the plaintiff and for the defendant, rather than as a separate category.

[13] Apart from settlement, the court may dismiss the case prior to trial without request of one of the parties. We drop these cases from the sample. In this paper we do not distinguish different forms of adjudication, such as court verdicts, jury verdicts and directed verdicts.

often when true case quality is near the court's decision standard, and this selection mechanism drives plaintiff win rates toward 50 percent. In AI models (Bebchuk, 1984), the probability that the plaintiff will win is private information. An uninformed party makes a settlement offer (or a sequence of offers, Spier, 1992), which is accepted by the informed party only when he has a low probability of winning at trial. Trials arise in (separating) equilibria because settlement offers have some probability of failing due to the information asymmetry. This one-sided selection mechanism predicts that the win rate for the informed party should tend toward 100 percent. We show that our data strongly favor the divergent expectations model for patent infringement suits.

3. Non-parametric Evidence

Table 1 presents estimates of average filing rates for each technology field during three sub-periods: 1978-84, 1985-90 and 1991-95. We measure filing rates as the number of suits filed per thousand patents from a given *cohort* (application year), and include multiple cases for the same patent.[14] Rates are based on suits filed through 1999, and then adjusted for truncation due to time lags in case filings and under-reporting.

Mean filing rates vary substantially across technology fields. Equality across fields is strongly rejected ($\chi^2(7) = 1{,}103$; p-value < 0.001). For the aggregate (pooled technology field) data, there are 19.0 case filings per thousand patents. The lowest rates are found in Chemicals (11.8), Electronics (15.4) and Mechanical (16.9). Interestingly, filing rates for pharmaceutical patents are only modestly higher than the average. The rates are much higher for patents in Other Health, Computers, Biotechnology, and Miscellaneous. Computers and Biotechnology are both newer areas where one might expect there to be greater uncertainty about legal outcomes.[15]

[14] We do not compute rates by case filing year because the population of patents alive in any year (the denominator) is unknown since it depends on patent renewals for the preceding cohorts. Further, the age structure of the population changes over time as patenting increases, and age and filing rates are related.

[15] Filing rates in Table1 include only the main patents in each suit, while there may in fact be several patents per suit. We present these calculations because, for filing years before 1990, we only have information about the main patents (mixing the subsidiary patents for later years would distort trends). "True" rates are even higher than those in Table 1 if one views being a subsidiary patent in a case as

The number of suits filed has been driven up sharply by increasing patenting in all technology fields and a shift of patenting toward those with higher litigation rates. The total number of patents grew by 71 percent over the period, but in Drugs, Biotechnology, and Medical Instruments patenting nearly tripled, and in Computers it grew four-fold. Once the growth in patenting is taken into account, the table demonstrates that there has been *no trend increase in the filing of suits* in any technology field over this period.

Table 2 summarizes mean filing and settlement rates for four ownership categories: individuals, domestic unlisted and listed companies, and foreign companies. With 10.4 suits per thousand patents, domestic listed companies are far less likely to prosecute infringements than unlisted companies and individuals with 35-45 suits. Moreover, filing rates for foreign patentees (mostly unlisted firms) are far lower than for their domestic counterparts. The joint null hypothesis that average filing rates are the same is decisively rejected (c^2 (3) = 11,853; p-value <0.001).

Although ownership type has a strong effect on filing rates, it does not affect the probability that a suit, once filed, is settled (c^2(3) = 4.55; p-value ≈ 0.2). About 95 percent of all patent suits are settled by the parties before the end of trial, and most of those before the trial begins, but systematic sorting on the ownership dimension occurs *before* suits are filed.

Listed firms may have lower filing rates for a number of reasons. Such firms have larger patent portfolios, which gives them both experience and the ability to settle disputes by pooling or trading intellectual property.[16] If imperfect capital markets limit the capacity of smaller firms to finance litigation, larger firms may be better able to settle (and to extract better terms) because they pose more credible litigation threats in confronting smaller firms. And when large firms have disputes with each other, they are likely to have many points of interaction, especially through competition in

equivalent to being the main litigated patent. One could scale up the filing rates using the ratio of subsidiary to main patents, which is 0.24 overall and for individual technology fields: Drugs 0.25, Other Health 0.36, Chemicals 0.20, Electronics 0.37, Mechanical 0.20, Computers 0.34, Biotechnology 0.46, and Miscellaneous 0.15.

[16] Patenting experience may facilitate settlement if it makes firms better able to assess the quality of their cases and likely outcomes in court. Experienced firms may also make higher quality patent applications that give rise to fewer disputes in the first place (Graham, *et. al.*, 2001).

product markets, which should promote settlement. The detailed patent data enable us to make progress in discriminating between these sources of advantage.

We begin by examining how the probability of litigation (i.e., of being involved in at least one suit over the life of the patent) and the probability of post-suit settlement vary with different portfolio sizes. To compute these probabilities, we adjust for the fact that patents from large portfolios are disproportionately represented in the matched data (since the matching was not stratified by portfolio size –Appendix 1 for details). The first column of Table 3 shows that the probability of litigation sharply declines with portfolio size. A formal test confirms this finding ($\chi^2(6) = 2,610$; p-value < 0.001). The probability of filing a suit involving a patent in a portfolio with a small number of other patents (0-10) is 1.7 percent, compared to about 0.5 percent for a patent in a portfolio with 100-300 other patents, and only 0.26 percent for those in large portfolios (> 900 patents). These are large differences, suggesting that having bigger portfolios confers substantial advantages.

If firms with larger portfolios have a higher propensity to patent their innovations, and thus more patents that are not worth fighting over, it could drive these differences. But the evidence contradicts this hypothesis. Portfolio size is *positively*, and significantly, correlated with forward citations and forward citations per claim. The correlation coefficients are 0.10 and 0.06, respectively (computed using the matched sample and cohorts 1978-1988 to avoid spurious correlation due to both portfolio size and citations being truncated). Even in electronics, where firms have been described as following a patent harvesting strategy, there is no evidence that the average quality of patents falls in larger portfolios. We conclude that the link between litigation probability and portfolio size reflects advantages that large portfolios give to firms in settling disputes.

To distinguish between the advantages of portfolio size and firm size, we separate unlisted, and small and large publicly listed domestic firms.[17] The litigation probability declines with portfolio size within each ownership type, at least in terms of the point estimates. However, it is by far more precipitous for domestic unlisted companies. For a patent owned by such a company and in a portfolio

[17] Listed firms that could not be classified because they lack employment data are not included in this table.

of 0-10 other patents, the *average* probability of being involved in litigation is 2.6 percent, while for patents in the same sized portfolio but owned by listed domestic companies it is less than one percent. At the same time, there is little evidence that size – either in terms of public listing or employment – matters once more than about 100 patents are held. These relationships hold in each of the technology fields (not reported).

The final column presents the average probability of settlement, conditional on a suit being filed, for different portfolio sizes. As in Table 2, we observe only small differences in the post-suit settlement rates. The differences across portfolio size are marginally statistically significant in the pooled data ($c^2(6) = 14.2$; p-value ≈ 0.05), and insignificant in tests controlling for ownership type (Lanjouw and Schankerman, 2001b).

To summarize, the likelihood of filing a suit (i.e., *not* settling beforehand) is much higher for patents owned by individuals and unlisted companies, and for patentees with smaller patent portfolios. But these differences do not appear in post-suit settlement rates. Thus almost all of the *sorting among patent disputes* occurs before suits are filed, not afterwards in the courts.

We now turn to the timing of such settlements and the win rates for cases that are adjudicated at trial. Table 4 presents this information by ownership type. About 80 percent of all suits that are ever settled (without third party adjudication) are settled before a pre-trial hearing is held. This suggests that the filing of a suit sends a strong signal about the seriousness of the plaintiff to use legal means, and quickly triggers resolution before substantial legal costs are incurred.[18] Nearly all of the remaining settlement occurs before the trial commences. As with the likelihood of settlement itself, the timing of settlements differs little by ownership type.

The table also shows the trial win rates (for infringement suits). Win rates are close to 50 percent, as predicted by the divergent expectations model of litigation. They are sharply inconsistent with the win rates of either zero or 100 percent predicted by asymmetric information models.

[18] Pooling all cases, the median period before settlement occurs (in months) is 8, 16, and 25 for those settling before pre-trial hearing, after a hearing but before trial, and after trial, respectively.

4. Econometric Analysis

In this section we present probit regressions of the probability of infringement suits and post-suit settlement for the pooled data.[19] We use the Derwent data as the basis for the sample and then include, in both estimations, only those cases linked to the FJC database and therefore having outcomes information. In analyzing the determinants of the *litigation probability* (filing of suits), we do not count multiple cases involving the same patent. We do this to avoid undue influence by a few patentees that might sue many infringers in separate but related cases. We include multiple cases in the econometric analysis of suit outcomes for three reasons: this is appropriate if the purpose is to assess litigation risk for pricing patent insurance; it is unclear how one would choose the "representative" suit when there are multiple cases; and the sample size for outcomes (especially trials) is relatively small even when we include multiple cases.

Table 5 contains variable definitions. Though the original matching controls for the effects of cohort and technology, we include technology group controls in the regressions because the samples used here differ somewhat. Table 6 presents the parameter estimates and the marginal effect of each variable on the *population* probability of litigation for a randomly drawn patent (at matched sample mean characteristics).[20] For comparison, the mean litigation probability is 0.0135. We first summarize the results for the control variables, and then focus on the main variables of interest.

The probability of litigation increases with the number of claims and forward citations per claim, and the effects are substantial. At means, the elasticity of the litigation probability with respect to the number of claims is 0.31. The fact that the elasticity is much less than one is interesting because it indicates that the standard procedure of pricing patent litigation insurance on a per claim basis is inappropriate. In contrast, the likelihood of a suit falls with the number of backward citations per claim. The effect is small, but the finding is consistent with the view that backward citations are an

---

[19] Results are similar for declaratory judgement suits (Lanjouw and Schankerman, 2001b).

[20] Since the sample litigation rate is close to 40 percent by construction, we deflate the sample marginal effects using a conversion factor to obtain those appropriate for a randomly drawn patent in the population (see Appendix 2). Marginal effects for ownership dummy variables include portfolio interactions. They are constructed so that the alternative to a given ownership category is a weighted

indication that the patent is in an already well-developed technology area where uncertainty about property rights is less likely to cause disputes.

The coefficient on the variable *Fwdself*, the percentage of citations that are self-citations, is positive and significant. By contrast, greater backward self-citation (*Bwdself*) significantly reduces the likelihood of litigation. This pattern is consistent with the hypothesis that self-citation indicates the presence of a "cumulative innovation" by the patentee, and that there is complementarity among technologically related patents in a firm's portfolio that raises the willingness to protect the property rights of the key, early inventions in the chain.

This first set of findings confirms the importance of patent value in determining litigation risk and related opposition proceedings within patent offices (see also Lanjouw and Schankerman, 2001a; Graham, *et.al.*, 2001; and Harhoff and Reitzig, 2000).

Using a sample of biotechnology patents, Lerner (1994) found that patents with uses in many technological areas – "broad" patents – are more likely to be litigated. In our more comprehensive and more recent data, a similar measure is also significant, but with the opposite sign. A ten percent increase in *No3USPC* (the mean is 2.2) reduces the litigation probability by about 1.7 percent.[21] It may be that it is harder to detect infringements when a patented innovation is used in more technology areas, and that this dominates any increase in the number of potential infringers.

We turn now to the core results. The most important is that the probability of litigation is negatively related to the size of the patent portfolio, with an elasticity (at means) of -0.13. The marginal effect declines with portfolio size, but the point estimate is negative over most of the sample range. Having a larger portfolio of patents reduces the probability of being involved in a suit on any individual patent.

---

average of the other categories, with weights reflecting their representation in the matched sample. Marginal effects for variables with quadratic terms are given as a composite.

[21] The point estimates in the separate technology fields (not reported) are negative and statistically significant in five cases, negative but insignificant in two, and positive but insignificant in one.

Being part of a large patent portfolio is much less important for drug patents.[22] The estimated coefficient on a dummy for non-drug technology fields (*PortNondrug*) is negative and large relative to the baseline portfolio effect. It implies that the marginal benefit of portfolio size on the litigation probability is nearly twice as large for non-drug patents as compared to drug patents (compare *Portsize* and *PortNondrug*). This finding is consistent with the idea that trading intellectual property is especially important in areas where innovation is "complex" in the sense that it may rely on multiple components or research tools that may be patented by other firms. This feature has been less important in the pharmaceutical industry.[23]

We allow for a set of interactions between portfolio size and firm ownership type (large domestic listed companies are the left out category). Controlling for other factors, the point estimates again strongly support the hypothesis that company size affects the importance of having larger patent portfolios. For a small domestic listed company (*DLIST-S*) with the mean portfolio size (1,420 patents), the marginal effect of portfolio size on the probability of litigation is about eight times larger than for a large listed company with the same portfolio (compare marginal effects for *Portsize* and *PortDLIST-S*). The marginal effect of portfolio size for small listed firms is even greater than that for unlisted firms.

Moreover, we easily reject the hypothesis that there are no ownership differences in the probability of involvement in an infringement suit when controlling for other factors (?^{2}(6) = 978.8, p-value < 0.001). The pattern of marginal effects points to five main findings. First, foreign individuals and unlisted (smaller) companies are much less likely to engage in infringement suits than their domestic counterparts. Comparing the marginal effects of *FIND* and *DIND*, the probability of litigation is much lower – by 1.2 percentage points – for foreign individuals than for their domestic counterparts. Comparing foreign and domestic unlisted companies (*FUNLIST* and *DUNLIST*), the

---

[22] Separate estimations by technology field (not reported) suggested this hypothesis. Other differences in estimated portfolio coefficients across technology were not statistically significant.

[23] Harhoff and Reitzig (2000) find that larger portfolios reduce the probability of opposition proceedings at the European Patent Office. Using related technology definitions, Somaya (2001) finds that portfolio size has a significant negative effect on U.S. litigation involving computer patents, but an insignificant effect for patents in research medicine.

difference is even greater – almost two percentage points. Second, larger domestic and foreign listed companies are equally likely to file suits. Third, domestic individuals, unlisted and small listed companies are equally likely to litigate (the differences in point estimates are not statistically significant). Fourth, domestic individuals and unlisted companies are more likely – by about one percentage point – to litigate than large domestic listed firms (*DLIST-L*). Finally, small listed companies are more likely to file suits than larger ones, again the difference being about one percentage point. To summarize, we find the following ranking: {*DLIST-S*, *DUNLIST*, *DIND*} > {*DLIST-L*, *FLIST*} > {*FIND*, *FUNLIST*}.

Since these effects are conditional on portfolio and company size (both of which relate to the cost of settling), the lower rate of litigation for foreign owners should reflect two main factors: the cost of pursuing U.S. court actions and access to information about potential infringements. We expect that the cost of litigating for domestic patentees is less than (or equal to) that for foreign patentees, and that it is harder for foreign owners to detect infringements in the United States. If so, domestic owners should litigate more often than their foreign counterparts and this is what we find, except for listed companies. This exception is not surprising, since foreign firms that are listed, with a presence in the United States, are less likely to be disadvantaged in terms of litigation costs and access to information.

The technology concentration variable, *C4*, provides additional evidence that the expectation of repeated interaction promotes settlement. If a company operates in concentrated area there is a greater chance that it will be involved in disputes with the same firms, allowing them to develop ways to resolve disputes cooperatively. As predicted, the estimated coefficient on *C4* is negative and highly significant. A ten percent increase in the index reduces the probability of a suit by 4.6 percent.[24]

The litigation probability is also influenced by the asymmetry in portfolio size between the patent owner and likely disputants (*Relsize*), which we interpret as reflecting the patentee's threat power. As expected, if a patent owner is large relative to typical disputants, the probability of a case

[24] Higher concentration could indicate a more mature technology with less valuable innovation. Thus it is important that the regressions control for patent value in order to distinguish between these interpretations.

filing is lower. However, the effect is not large – a ten percent increase in relative size lowers the litigation probability by just 0.5 percent.

Table 7 highlights the huge heterogeneity in litigation risk implied by these estimation results. We calculate the probability of an infringement suit for each patent in the matched sample, given the patent's full set of characteristics. Percentile cutoffs for the distribution of these probabilities are given in the first row of the table. The probability of litigation for the median patent is just under one percent. However, among the top one percent of patents (99th percentile), the probability of involvement in a suit is over eight percent. The table shows that the rates can be far higher when patents are segregated into different technology and ownership groups. The top percentile of patents in areas that are most at risk have probabilities of litigation over 15 percent (see Other Health, Computers and Biotechnology). Similarly, the top one percent of all patents held by domestic unlisted firms have litigation risks of well over 10 percent. Since evidence from patent renewal data and firm surveys indicates that private value of innovations is highly skewed – with most value attributable to the top patents – it is precisely the litigation risk in these top percentiles that is relevant for how enforcement costs affect R&D incentives.

Finally, we turn to the econometric analysis of several post-suit outcomes. In estimating these regressions, we do not control for selection. Rather, we ask: given any selection that occurs, is there any remaining association between patent and patentee characteristics and the outcomes? For purposes of assessing *ex ante* litigation risk (e.g., for patenting decisions or insurance pricing), this is the relevant question.[25]

The evidence presented in the previous section indicated that the main characteristics of patents and their owners do not affect the probability of settlement after a suit is filed, nor the plaintiff win rates for cases that reach trial. Probit regressions for settlement and win rates confirm this conclusion. For brevity we summarize the findings but do not present the parameter estimates. The settlement regression has a meager pseudo-R^2 of 0.01. The null hypothesis that the regression as a whole is insignificant is not rejected (c^2(29) = 39.7; p-value = 0.089). The only positive finding is that the coefficients on three technology field dummies are significant and indicate that the settlement probability is about eight percentage points higher for patents in Electronics, Mechanical and Miscellaneous.[26] The probit regression for win rates has a pseudo-R^2 of 0.02 and is statistically insignificant (c^2(28) = 19.7; p-value = 0.90), as are the individual coefficients. Based on our discussions with staff at the FJC, there is no reason to believe that the data on settlements and plaintiff win rates are systematically bad (these outcome data are recorded at different times and in many different courts). We are confident that the "insignificance" of these regressions is meaningful: i.e., settlement and win rate outcomes are almost completely independent of observed characteristics of patents and their owners.

The probability that the settlement of infringement suits occurs early (before the pre-trial hearing) is also unrelated to most characteristics of the patent and its owner, with three noteworthy exceptions (the probit is significant: c^2(29) = 50.5, p-value = 0.008). First, early settlement is more likely if the patent in dispute is part of a larger portfolio. This is consistent with our earlier result that portfolio size makes filing a suit less likely in the first place. Second, patent owners that are large relative to a representative disputant (*Relsize*) are less likely to settle early. Greater technology concentration (*C4*) also makes early settlement somewhat less likely. Recall that the probability of a suit filing is lower when the relative size of the patentee is larger, which we interpret as reflecting greater threat power. But if the (implicit) threats do not succeed in preventing the need to file suit, it may be important for the patentee to carry out those threats in order to maintain credibility (post-suit "toughness"). Similarly, if the discipline of repeated interaction has failed to keep firms in a

---

[25] Controlling for filing selection would be appropriate if one wanted to infer the effect of characteristics on outcomes should a random sample of patents be subjected to suits. Evidence that there is a sample selection effect is mixed (Somaya 2001).

[26] It is also interesting to note that, if we restrict attention to suits where the patentee is identified as the plaintiff, then those suits involving smaller patentees (unlisted firms and domestic individuals) are significantly less likely to settle. These are patentees who do *not* have an exclusive licensee or late assignee litigating in their place. As plaintiffs they are more likely to be inexperienced and more attached to their innovations than owners who have licensed or sold out. Both characteristics could impede settlement.

concentrated area out of court in the first place, the dispute is probably very intractable. Both could delay any post-suit settlement, and this is what we find.

5. Implications for Pricing Patent Litigation Insurance

Our findings that patent portfolio size and technology concentration significantly affect litigation risk have important implications for R&D incentives. The threat of costly enforcement can affect R&D investment and patenting strategies (Waterson, 1990; Lerner, 1995). This is especially so for small, high technology firms that are more likely to face capital market constraints (Himmelfarb and Petersen, 1994). Shapiro (2001) emphasizes that firms rely heavily on cross-licensing arrangements and patent pools as a way of mitigating these problems of the anti-commons (fragmented property rights). But small firms are effectively blocked from using these arrangements unless cash payments are accepted for participation, and typically they are not.

One potential alternative for small firms is to rely on patent litigation insurance. This has been available in the United States and Europe for more than a decade, but the market remains severely under-developed.[27] The central problem is that the level and structure of pricing need to be rationalized. The level of prices is very high – Kaplan (2001) reports that, as a rule of thumb, the *annual* premium for an insured of average risk is about 6-7 percent of the coverage. This is much higher than the litigation risk we estimate, even for the most exposed patents. In our data the average probability of engaging in at least one infringement suit over the *entire lifetime* of a patent is 1.35 percent; rising to eight percent for the 99th percentile (Table 7). Including also suits for declaratory judgement, the total number of expected cases involving a given patent, over its lifetime, is just 1.9 per hundred patents (Table 1). Second, insurers typically set the premium on the basis of the coverage per claim (which is normally the same as the maximum aggregate coverage allowed). But as we showed earlier, the litigation probability varies much less than proportionally to the number of claims, so patents with more claims are being 'overpriced.' Finally, the huge heterogeneity in litigation risk, illustrated in Table 7, is not systematically reflected in insurance premiums. Insurers make some attempt to differentiate prices on the basis of detailed examination of patents, but this is extremely costly. The empirical work in this paper provides insurance companies with a useful starting point for rationalizing prices in a more systematic, and less expensive, way and thereby for mitigating adverse selection.[28]

As protection for insurers, policies often have payback provisions under which the insured party absorbs (part of) the litigation costs from monetary payoffs from settlement or trial, and in some cases the insurer shares in those payoffs. This raises a second key problem, which is how to assign control rights over when to accept settlement offers. In most policies we have examined, the insurer retains some measure of control rights, but these rights are often vaguely specified and expose patent holders to a form of *ex post* hold-up.

6. Concluding Remarks

We analyze the determinants of patent infringement suits, and their outcomes, by linking detailed information from the U.S. patent office to data from the U.S. federal court system, the Derwent database and industry sources. The linked data allow us to construct a suitable controlled random sample of the population of potential disputants. The data set we construct is the most comprehensive yet available, covering all patent suits in the United States reported by the federal courts during the period 1978-1999.

A major finding in the paper is that all of the sorting among patent disputes on observed characteristics occurs in the decision to file suits. The key post-suit outcomes do not depend on these characteristics. From a policy perspective, this is good news because it means that enforcement of patent rights relies on the effective *threat* of court action more than on extensive legal proceedings.

[27] For discussion of patent litigation insurance experience, see Danish Ministry of Trade and Industry (2001). See also Kaplan (2001). Lloyds of London, for example, offered patent litigation insurance in the 1980's but discontinued the policy due to lack of profitability.
[28] Most of the variables in the analysis are completely, or largely, out of the control of the patentee. Before linking insurance prices to observable characteristics, consideration would need to be given to the patentee's ability to manipulate them.

This feature is reinforced by high post-suit settlement rates and the fact that most settlement occurs quickly, often before a pre-trial hearing is held. These findings mean that the enforcement of patent rights minimizes the use of judicial resources for resolving disputes. The bad news is that individuals and small companies are much more likely to be involved in suits conditional on the characteristics of their patents, and this is not offset by a more rapid resolution of their suits.We also show that patentees are better able to avoid going to court when they have a portfolio of intellectual property, orwhere there are other features of the environment that promote "cooperative" behavior. In this sense, small firms are at a disadvantage in protecting their intellectual property. But the fact that heterogeneity in litigation risk is measurable offers the prospect of developing a market for more affordable patent litigation insurance.

An important direction for future research is to explore the dynamic aspects of conflict between firms over intellectual property assets. This would include studying multiple (sequential) suits on the same patent between different parties, and multiple suits on different patents involving the same parties. Initial work along these lines for a sample of cases is found in Somaya (2001). To make further progress, we are undertaking a project to match litigants across all patent cases. When completed, these data will provide information about the role of reputation building in settlement negotiations. In addition one could use our empirical estimates of litigation risk, together with information on legal costs at different stages of the litigation process, to calculate the *ex ante*, direct costs of enforcing patent rights. Theoretical work is needed to understand how patent litigation insurance markets affects strategic interaction and settlement outcomes among firms. This would enable us to measure the indirect costs of enforcing patent rights, and hence the total effect of enforcement on R&D incentives.

References

American Intellectual Property Law Association (AIPLA), 2001, *Report of the Economic Survey*, Washington, D.C.

Bebchuk, Lucian, 1984, "Litigation and Settlement under Imperfect Information," *RAND Journal of Economics*, vol. 15, 404-415.

Berheim, Douglas and Michael Whinston, 1990, "Multimarket Contact and Collusive Behavior," *The RAND Journal of Economics*, 21(1), 1-26.

Cooter, Robert and Daniel Rubinfeld, 1989, "Economic Analysis of Legal Disputes and their Resolution," *Journal of Economic Literature*, vol. 27, 1067-97.

Danish Ministry of Trade and Industry, 2001, *Economic Consequences of Legal Expense Insurance for Patents*. Report prepared for the Danish Patent Office by the Economic Analysis Group. Copenhagen.

Eisenberg, Rebecca, 1999, "Patents and the Progress of Science: Exclusive Rights and Experimental Use," *University of Chicago Law Review*, vol. 56, 1017-55.

Federal Judicial Center, Federal Court Cases: Integrated Data Base, 1970-89. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Tapes updated to1999.

Graham, Stuart, Bronwyn Hall, Dietmar Harhoff and David Mowery, 2001, "Post-Issue Patent 'Quality Control': A Comparative Study of U.S. Patent Re-examinations and European Patent Oppositions," University of California at Berkeley. Mimeo.

Grindley, Peter and David Teece, 1997, "Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics," *California Management Review,* 39(2), 8-41.

Hall, Bronwyn and Rosemarie Zeidonis, 2001, "The Patent Paradox Revisted: An Empirical Study of Patenting in the Semiconductor Industry, 1979-1999," *RAND Journal of Economics*, 32(1), 101-128.

Harhoff, Dietmar and Markus Reitzig, 2000, "Determinants of Opposition against EPO Patent Grants - The Case of Biotechnology and Pharmaceuticals," Muenchen: Ludwig-Maximilians-Universitaet. Mimeo.

Heller, Michael and Rebecca Eisenberg, 1998, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science* , 698-701.

Himmelfarb, Charles and Bruce Petersen, 1994, "R&D and Internal Finance: A Panel Study of Small Firms in High-Technology Industries," *Review of Economics and Statistics*, vol. 76, 38-51.

Jaffe, Adam and Manuel Trajtenberg, 1999, "International Knowledge Flows: Evidence from Patent Citations," *Economics of Innovation and New Technology*, vol. 8, 105-136.

Kaplan, Jonathan, 2001, "Patent Litigation Insurance: An Introduction," *Patent Strategy and Management*, 1(9), 1-8.

Lanjouw, Jean O. and Josh Lerner, 2001, "Tilting the Table? The Predatory Use of Preliminary Injunctions," *The Journal of Law and Economics*, 44(2), 573-603.

Lanjouw, Jean O., Ariel Pakes and Jonathan Putnam, 1998, "How to Count Patents and Value Intellectual Property: Uses of Patent Renewal and Application Data," *Journal of Industrial Economics*, vol. XX (December), 405-32.

Lanjouw, Jean O. and Mark Schankerman, 2001a, "Characteristics of Patent Litigation: A Window on Competition," *RAND Journal of Economics*, 32(1), 129-151.

Lanjouw, Jean O. and Mark Schankerman, 2001b, "Enforcing Intellectual Property Rights," NBER Working Paper no. 8656 (December).

Lerner, Joshua, 1994, "The Importance of Patent Scope: An Empirical Analysis," *RAND Journal of Economics*, vol. 25, 319-333.

Lerner, Joshua, 1995, "Patenting in the Shadow of Competitors," *Journal of Law and Economics*, vol. 38, 463-96.

P'ng, I.P.L., 1983, "Strategic Behavior in Suit, Settlement and Trial," *Bell Journal of Economics*, vol. 14, 539-550.

Priest, G. and B. Klein, 1984, "The Selection of Disputes for Litigation," *Journal of Legal Studies*, vol.13, 1-55.

Shapiro, Carl, 2001, "Navigating the Patent Thicket: Cross Licenses, Patent Pools and Standard-Setting," in Adam Jaffe, Joshua Lerner and Scott Stern, eds., *Innovation Policy and the Economy* (Cambridge: MIT Press for the NBER), vol. 1, 119-50.

Siegelman, Peter and Joel Waldfogel, 1999, "Toward a Taxonomy of Disputes: New Evidence Through the Prism of the Priest/Klein Model," *Journal of Legal Studies*, 18(1), 101-130.

Somaya, Deepak, 2001, "My Strategy Says: 'See You in Court!' Determinants of Decisions not to Settle Patent Litigation in Computers and Research Medicines," Robert Smith School of Business, University of Maryland. Mimeo.

Spier, Kathryn, 1992, "The Dynamics of Pretrial Negotiation," *Review of Economic Studies,* 59(1), 93-108.

Tirole, Jean, 1994, *The Theory of Industrial Organization* (Cambridge, MA: MIT Press).

Waldfogel, Joel, 1998, "Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation," *Journal of Law and Economics*, XLI (October), 451-476.

Waterson, Michael, 1990, "The Economics of Product Patents," *American Economic Review*, 80(4), 860-869.

Appendix 1. Computing Population Filing Probabilities by Portfolio Size, and their Variance

Let L_{gz}, M_{gz} and N_{gz} denote, respectively, the number of patents in the litigated and matched samples and in the population that are in portfolios of size z and from group g, where the latter is defined by technology field, cohort, and ownership type. True filing probabilities in the population are p^*_{gz}, which can be estimated by p_{gz} = [L_{gz} / N_{gz}]. We cannot calculate p_{gz} directly as N_{gz} is unobserved. However, since the matched sample is random with respect to portfolio size, we can use the sample share of the patents in group g that are in portfolios of size z, $\hat{s}_{gz}$ = [M_{gz}/ M_g], as an unbiased estimator of the population share [N_{gz}/ N_g]. Using this, our estimator is:

$$\hat{p}_{gz} = \frac{L_{gz}}{N_g}\frac{1}{\hat{s}_{gz}}.$$

[L_{gz} / N_g] = $\hat{p}_{gz}$ is also an estimate of an underlying probability, say p_{gz}, with an associated sampling variance. Taking a Taylor expansion, we can capture both sources of error in the following approximation:

$$Var(\hat{p}_{gz}) = Var\left(\hat{p}_{gz}\frac{1}{\hat{s}_{gz}}\right) \approx \left[\frac{-\hat{p}_{gz}}{\hat{s}^2_{gz}}\right]^2 \frac{\hat{s}_{gz}(1-\hat{s}_{gz})}{M_g} + \left[\frac{1}{\hat{s}_{gz}}\right]^2 \frac{\hat{p}_{gz}(1-\hat{p}_{gz})}{N_g},$$

where the covariance terms are zero because the two sources of sampling error are independent. This simplifies to:

$$Var(\hat{p}_{gz}) = \hat{p}^2_{gz}\left[\frac{(1-\hat{s}_{gz})}{M_{gz}}\right] + \frac{\hat{p}_{gz}}{N_g}\left[\frac{1}{\hat{s}_{gz}} - \hat{p}_{gz}\right].$$

Filing probabilities at a more aggregated level are calculated as a weighted average of these rates, with weights based on M_g.

Appendix 2. Deriving Population Litigation Probabilities and Marginal Effects

2.1 Population Litigation Probabilities

We define classes by characteristics with respect to which the sampling was non-random: technology groups, cohort, infringement suits, and declaratory judgement suits. Let $P(X_c)$ denote the population probability of litigation for a patent in class c with a vector of other characteristics X_c and let $Q(X_c)$ be the corresponding probability in the pooled (litigated and matched) sample. $P(X_c)$ and $Q(X_c)$ differ because the matched sample was constructed so that the overall litigation probability is fifty percent, controlling for technology and cohort. We want to infer $P(X_c)$ from the estimated value of $Q(X_c)$.

First we determine the extent to which we must inflate the matched sample for a given class to have it reflect the number of patents in that class in the population. Let Q represent the aggregate sample litigation probability for a given class:

$$Q = L / (L + M),$$

where L and M denote the number of litigated and matched patents in the sample. The population probability is

$$P = L / N,$$

where the number of litigated patents is the same in both cases since the sample contains all (reported) litigated patents, and N is the total number of patents in the class in the population. Thus,

$$N = \{Q/(1-Q)P\}M \equiv KM$$

Within a class, the matched patents are random draws so the distribution of characteristics in the matched sample is the same as the population. Thus the expected number of matched patents with characteristics X_c in the population, $N(X_c)$, is greater than in the sample by the inflation factor, K, and so equals $KM(X_c)$. Letting $L(X_c)$ be the number of litigated patents with characteristics X_c, the expected population probability of litigation for such patents is

$$P(X_c) = L(X_c)/[KM(X_c)].$$

Similarly, $Q(X_c) = L(X_c)/[L(X_c) + M(X_c)]$. Solving for M and substituting, we get the result:

$$P(X_c) = Q(X_c)/[K(1-Q(X_c))] \qquad \text{(A.2.1)}$$

2.2 Population Marginal Effects

For each characteristic X_k, the population marginal effect is

$$\partial P(X_c)/\partial X_{kc} = [dP(X_c)/dQ(X_c)]\ \partial Q(X_c)/\partial X_{kc}$$

The last term is the sample marginal effect computed from the probit regression. From the expression for $P(X_c)$ we get

$$dP(X_c)/dQ(X_c) = 1/K[1\text{-}Q(X_c)]^2$$

Measuring $Q(X_c)$ by the sample probability of litigation in the class, Q, we get the result:

$$dP(X_c)/dQ(X_c) \approx P/Q(1\text{-}Q)$$

We measure *P* for each class as follows: For the denominator, we take the total number of patents is the class during 1978-1995. In the numerator we use the number of infringement suits that can be directly identified, and all unidentified suits. These are inflated for under-reporting and truncation. We then calc ulate marginal adjustment factors by USPC groups. Separate classes defined by cohort are not used because of our finding that litigation probabilities have been similar over time within technology areas. Because $dP(X_c)/dQ(X_c)$ is the same for all X_k for a given class *c*, all sample marginal effects are adjusted by the same factor to convert them to population marginals. The aggregate conversion factor is used in Table 6.

Conversion Factors to Estimate Population Marginal Effects

Aggregate	.048
Drugs	.050
Other Health	.089
Chemicals	.031
Electronics	.038
Mechanical	.045
Computers	.063
Biotechnology	.076
Miscellaneous	.084

Table 1. Filing Rates by Technology Fields and Cohort Groups

	Filing Rate (cases per thousand)			
	Total: *1978-95*	*1978-84*	*1985-90*	*1991-1995*
Technology Field (percent of sample)				
Aggregate	**19.0**	**19.3**	**16.6**	**21.1**
(100.0)	(0.21)	(0.31)	(0.28)	(0.44)
Drugs	**22.2**	**22.5**	**18.9**	**24.3**
(5.6)	(1.05)	(1.62)	(1.34)	(1.97)
Other Health	**34.6**	**48.2**	**35.2**	**27.3**
(8.0)	(1.33)	(2.67)	(1.98)	(2.23)
Chemicals	**11.8**	**11.6**	**10.9**	**13.0**
(13.4)	(0.35)	(0.50)	(0.49)	(0.80)
Electronics	**15.4**	**16.2**	**13.1**	**16.8**
(18.7)	(0.40)	(0.61)	(0.51)	(0.79)
Mechanical	**16.9**	**17.7**	**14.5**	**18.7**
(27.7)	(0.2)	(0.53)	(0.46)	(0.79)
Computers	**25.6**	**32.6**	**21.2**	**25.9**
(1.8)	(2.25)	(4.24)	(2.80)	(3.78)
Biotechnology	**27.9**	**33.3**	**27.6**	**25.5**
(0.9)	(3.36)	(6.13)	(5.16)	(5.52)
Miscellaneous	**34.2**	**32.4**	**28.9**	**40.7**
(23.9)	(0.76)	(1.10)	(0.98)	(1.66)

Notes: Percent of sample is the breakdown of litigated patents from cohorts 1978-95. Estimated filing rates including multiple suits involving the same patent. Figures use data for suits filed through 1999, adjusted for under-reporting and truncation. Estimated standard errors are in parentheses.

Table 2. Filing and Settlement Rates, by Ownership Type

	Individuals	*Domestic Unlisted*	*Domestic Listed*	*Foreign Firms*
Filing Rate	**35.2**	**46.0**	**10.4**	**4.2**
(cases/thousand)	(0.65)	(0.78)	(0.27)	(0.16)
Settlement Rate	**94.7**	**94.0**	**94.1**	**94.5**
(percent)	(1.4)	(0.4)	(0.7)	(0.9)

Notes: Foreign firms include listed and unlisted companies. The filing rate is the estimated number of suits per thousand patents from cohorts 1978-1995 (as in Table 1). The settlement rate is the fraction of filed cases reported to have been settled at some time prior to court judgment, according to the FJC. Settlement rates are computed for suits filed during the period 1978-1992 to minimize truncation bias and include only infringement suits. Estimated standard errors are in parentheses.

Table 3. Probabilities of Litigation and Settlement

Probability of:				**Litigation**					**Settlement**	
	Aggregate		*Large Domestic Listed*		*Small Domestic Listed*		*Domestic Unlisted*		**Aggregate**	
Portfolio Size										
0-10	**1.71**	(0.05)	**0.55**	(0.26)	**1.09**	(0.49)	**2.63**	(0.09)	**95.0**	(0.5)
11-100	**1.20**	(0.05)	**1.16**	(0.25)	**1.78**	(0.32)	**2.00**	(0.09)	**93.3**	(0.7)
101-200	**0.52**	(0.05)	**0.70**	(0.14)	**0.77**	(0.28)	**0.67**	(0.12)	**93.0**	(1.7)
201-300	**0.43**	(0.06)	**0.49**	(0.17)	**0.82**	(0.32)	**0.84**	(0.27)	**97.0**	(1.3)
301-600	**0.39**	(0.04)	**0.54**	(0.10)	**0.70**	(0.31)	**0.56**	(0.10)	**90.9**	(1.9)
601-900	**0.34**	(0.04)	**0.62**	(0.10)	**0.44**	(0.25)	**0.34**	(0.12)	**93.3**	(2.5)
>900	**0.26**	(0.01)	**0.39**	(0.02)	Nc		**0.37**	(0.06)	**93.2**	(1.1)

Notes: The probability of litigation is the estimated number of patents involved in suits (multiple suits *not* counted) per hundred patents, adjusted for under-reporting and truncation and for the over-representation of patents from large portfolios (Appendix 1). 'nc' denotes an empty cell. See also notes to Table 2.

Table 4. Timing of Settlement and Trial Win Rates, by Ownership Type

Timing of settlement (%)	*Domestic Listed*	*Domestic Unlisted*	*Foreign Firms*	*Individuals*
Before pre-trial hearing	**81.2**	**83.0**	**78.8**	**84.7**
	(1.2)	(0.7)	(1.7)	(0.8)
Before Trial	**18.0**	**15.5**	**19.9**	**14.2**
	(1.2)	(0.7)	(1.7)	(1.8)
During Trial	**0.8**	**1.5**	**1.3**	**1.1**
	(0.3)	(0.2)	(0.5)	(1.9)
Plaintiff win rate at trial	**51.2**	**49.1**	**42.7**	**46.5**
	(3.8)	(2.3)	(4.9)	(2.3)

Notes: The timing of settlements is computed on the basis of all infringement cases filed during the period 1978-1992 and terminated by settlement before or during trial, according to the FJC. Cases that proceed beyond trial (e.g., on appeal or remand, which are about 5%) are not included. The plaintiff win rate is the number of infringement cases where the court judgment favors the patentee divided by the total number of cases. When the FJC reports a judgment in favor of both parties, we treat it as a win for each party and adjust the total appropriately. Estimated standard errors in parentheses are based on the binomial formula.

Table 5. Variable Definitions	
Claims	Number of claims in the patent specification
Fwd cites/claim	Number of citations to the patent by subsequent patents, divided by claims.
Bwd cites/claim	Number of citations to prior patents in the patent specification, divided by claims.
Fwdself	Percentage of forward citations that are from patents owned by the same company code. For individuals it is set to zero.
Bwdself	Percentage of backward citations that are to patents owned by the same company. For individuals it is set to zero.
No3USPC	Number of unique three-digit technology classes to which the patent is assigned by the patent office examiner.
Portsize	Number of other patents owned by the same assignee that have an application year within a ten-year window of the application year of the patent in question. For individuals it is set to one.
PortNondrug	Portsize times an indicator variable that is one if the patent is not a drug innovation, zero if it is a drug innovation.
PortUNLIST	Portsize times UNLIST (see below)
PortFLIST	Portsize times FLIST (see below)
PortDLIST-S	Portsize times DLIST-S (see below)
Tech. Concentration (*C4*)	*C4* concentration of patenting activity – weighted average over the technology areas of the patent's forward citations.
Relsize	Total portfolio size of the patent owner divided by a weighted average of portfolio sizes of firms in the technology areas of the patent's forward citations.
FIND	Foreign (non-U.S.) individual
DIND	Domestic (U.S.) individual
FUNLIST	Foreign company assignee without a Standard & Poor's CUSIP code
DUNLIST	Domestic company assignee without an S&P CUSIP code
FLIST	Foreign publicly listed company (with an S&P CUSIP code)
DLIST-S	Domestic publicly listed company with < the median number of employees for such firms (5,425)
DLIST-L	Domestic publicly listed company with ≥ the median number of employees.

Table 6. Probit Estimation of Litigation Probability for Infringement Suits

Variable	**Parameter**	**Marginal**	**Variable**	**Parameter**	**Marginal**
Claims	**0.023**	0.00034	FIND	**-0.54**	-0.0058
	(0.001)			(0.09)	
Claims^2 (x 10^3)	**-0.024**		DIND	0.13	0.0067
	(0.002)			(0.08)	
Fwd cites/claim	**0.19**	0.028	FUNLIST	**-0.69**	-0.0106
	(0.008)			(0.08)	
$[\text{Fwd cites/claim}]^2$ (x 10^3)	**-4.38**		DUNLIST	**0.21**	0.0091
	(0.32)			(0.08)	
Bwd cites/claim	**-0.056**	-0.00082	FLIST	-0.15	0.0003
	(0.010)			(0.19)	
$[\text{Bwd cites/claim}]^2$ (x 10^3)	**0.89**		DLIST-S	**0.27**	0.0082
	(0.43)			(0.11)	
Fwdself	**0.51**	0.0082	DLIST-L	**-0.23**	-0.0014
	(0.058)			(0.08)	
Bwdself	**-0.31**	-0.0048	Tech. Concentration (*C4*)	**-4.17**	-0.049
	(0.08)			(.23)	
No3USPC	**-0.068**	-0.0011	Relsize (x 10^3)	**-3.1**	-0.065
	(.008)			(1.12)	
Portsize (x 10^3)	**-0.104**	-0.0012			
	(0.037)				
Portsize^2 (x 10^6)	**0.009**				
	(0.001)				
PortNondrug (x 10^3)	**-0.061**	-0.001			
	(0.033)		No. Observations	17,443	
PortUNLIST (x 10^3)	**-0.027**	-0.00043	Pseudo-R^2	0.162	
	(0.013)		c^2	3858.3	
PortFLIST (x 10^3)	0.001	0.0001			
	(0.020)				
PortDLIST-S (x 10^3)	**-0.6**	-0.0096			
	(.27)				

Notes: Estimated standard errors in parentheses. Numbers in bold are significant at the 0.01 level. Population marginals calculated as described in Appendix 2.

Table 7. Predicted Probabilities of Infringement Suits

Percentile of Distribution:	**99th**	**95th**	**50th**
Aggregate	**7.9%**	3.8%	0.8%
Technology Field			
Drugs	**9.4%**	3.9%	0.9%
Other Health	**19.5**	6.1	1.7
Chemicals	**4.2**	2.1	0.5
Electronics	**7.1**	2.8	0.5
Mechanical	**6.5**	2.8	0.7
Computers	**14.8**	4.5	0.6
Biotechnology	**12.9**	6.3	1.3
Miscellaneous	**8.3**	4.6	1.9
Ownership Type and Portfolio			
Domestic Individual	**9.4%**	4.4%	1.9%
Domestic Unlisted	**13.7**	5.9	1.9
Portfolio £ 600 patents	*14.4*	*6.3*	*2.1*
> 600	*6.8*	*2.4*	*0.7*
Small Domestic Listed	**6.3**	5.4	1.8
Portfolio £ 600 patents	*6.3*	*5.4*	*2.0*
> 600	*3.5*	*3.5*	*0.8*
Large Domestic Listed	**4.8**	2.0	0.5
Portfolio £ 600 patents	*5.4*	*2.6*	*0.9*
> 600	*4.1*	*1.5*	*0.4*
Foreign Listed	**2.5**	1.4	0.3
Foreign Individual	**4.2**	1.4	0.6
Foreign Unlisted	**1.4**	0.8	0.3

Note: The distributions are calculated by computing sample probabilities using the parameter estimates in Table 6, and then deflating to population probabilities using Appendix equation (A.2.1).