

NBER WORKING PAPER SERIES

THE LITIGIOUS PLAINTIFF HYPOTHESIS:
CASE SELECTION AND RESOLUTION

Theodore Eisenberg
Henry S. Farber

Working Paper 5649

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1996

Much of the work on this study was done while Professor Farber was a John M. Olin Fellow at Cornell Law School. The authors thank Joanne Gowa, James Powell, two anonymous referees, and attendees at conferences at Berkeley, Princeton, Cornell, U.C. Santa Cruz, Harvard, Michigan, Stanford, Yale, and The Rand Corporation for comments on an earlier draft. This paper is part of NBER's research programs in Law and Economics, and Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1996 by Theodore Eisenberg and Henry S. Farber. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

THE LITIGIOUS PLAINTIFF HYPOTHESIS:
CASE SELECTION AND RESOLUTION

ABSTRACT

A central feature of the litigation process that affects case outcomes is the selection of cases for litigation. In this study, we present a theoretical framework for understanding the operation of this suit selection process and its relationship to the underlying distribution of potential claims and claimants. We implement the model empirically by assuming that individuals vary more in their litigiousness (inverse costs of litigation) than do corporations. This assumption, coupled with the case selection process we present, yields clear predictions on trial rates as a function of whether the plaintiff and defendant were individuals or corporations. The model also yields a prediction on the plaintiff's win rate in lawsuits as a function of the plaintiff's identity. Our empirical analysis, using data on over 200,000 federal civil litigations, yields results that are generally consistent with the theory. Lawsuits where the plaintiff is an individual are found to have higher trial rates and lower plaintiff win rates.

Theodore Eisenberg
Cornell School of Law
Myron Taylor Hall
Ithaca, NY 14850

Henry S. Farber
Industrial Relations Section
Firestone Library
Princeton University
Princeton, NJ 08544
and NBER

The Litigious Plaintiff Hypothesis: Case Selection and Resolution

Theodore Eisenberg and Henry S. Farber

1. Introduction

A rich literature covers many aspects of suit, settlement, and trial of civil cases.¹ Yet almost none of the literature investigates the influence of the process by which plaintiffs decide whether to file suit on subsequent events such as trial rates and case outcomes.² This neglect of the first stage in the suit-settlement-trial sequence is problematical because any complete model of the post-filing process ultimately must take account of the pre-filing selection mechanism. This study demonstrates the importance of that mechanism.

We argue that the process through which potential claims either result or do not result in lawsuits has important influences on expected trial rates and suit outcomes. We present a framework for analyzing this suit selection process, and we develop specific implications for trial rates and plaintiff win rates. Then we examine data on over 200,000 civil suits filed in federal courts between 1986 and 1994, and we find results that are fully consistent with the implications of the model and establish the need to account for pre-filing selection processes in explaining post-filing case outcomes.

We begin with a startling but empirically sound premise: people are not litigious. Studies of medical malpractice victims and general household surveys reveal a surprisingly nonlitigious society. For example, fewer than ten percent of malpractice and products liability victims initiate legal action (Danzon, 1985; Report of the Harvard Medical Practice

Much of the work on this study was done while Professor Farber was a John M. Olin Fellow at Cornell Law School. The authors thank Joanne Gowa, James Powell, two anonymous referees, and attendees at conferences at Berkeley, Princeton, Cornell, U. C. Santa Cruz, Harvard, Michigan, Stanford, Yale, and The Rand Corporation for comments on an earlier draft.

¹ Useful surveys and citations may be found in, e.g., Cooter & Rubinfeld, 1989; Hay, 1995; Miller, 1994).

² Priest & Klein's (1984) analysis of the selection of cases for litigation analyzed which of the lawsuits filed eventuate in trials, and they derived some clear predictions regarding case outcomes. The central prediction in the limiting cases is that cases for trial will be selected such that the plaintiff and defendant win rates will each be 0.5 in tried cases regardless of the average underlying merits of the overall set of suits filed. Waldfogel (1995) extends this analysis and shows that there can be systematic deviations from the Priest-Klein result. Eisenberg (1990) shows that such deviations exist.

Study, 1990; Hensler et al., 1991).³ One explanation for this fact is that individuals vary substantially in how costly they find litigation and that only those individuals with low costs of litigation file a lawsuit when they feel they have been wronged or harmed. These costs are interpreted broadly, and they include not only pecuniary costs and costs due to risk aversion but also the psychological and emotional costs of confrontation. Litigation is confrontational, emotionally draining, uncertain, and a distraction from daily life. Litigation requires retaining an agent to manage one's affairs through a complex system incomprehensible to most laymen. It is almost always an extraordinary event in the lives of the parties. Not surprisingly, most people dislike litigation and avoid engaging in it even when there is reason to do so.

In this context, it is clear that case selection, the process through which potential claims are translated into lawsuits, depends not only on the monetary expected value of the claim but also on the non-pecuniary costs of litigation (the inverse of the taste for litigiousness) of the potential claimant. The central theme of our analysis is that cases are selected for suit systematically on the basis of the tastes for litigiousness of the plaintiffs so that the set of plaintiffs in actual lawsuits are a random draw of neither the general population nor the population of potential claims. Plaintiffs are drawn from those individuals with the highest taste for litigation (lowest costs of litigation) conditional on the expected value of the claim.

In the next section, we discuss from a theoretical perspective how this selection from the lower tail of the plaintiff's cost distribution can have important implications for the outcomes of litigation. Specifically, we show that, for a fairly general class of distributions of litigation costs, increasing the dispersion of litigation costs among potential plaintiffs results in a lower average litigation costs among plaintiffs who actually file lawsuits. We then argue that a property of any reasonable model of the litigation process is that lower litigation costs will imply higher trial rates. On this basis, we conclude that cases filed from pools of potential plaintiffs with greater dispersion in the distribution of their litigation costs will have a higher trial rates. We further argue that reasonable models of the litigation

³ The rate of claiming by victims of automobile accidents is distinctively higher than the rate of claiming for other tort victims (Hensler et al., 1991).

process will have the property that plaintiffs with lower litigation costs will be willing to file cases in which they have a smaller probability of prevailing at trial. Thus, we also conclude that cases filed from pools of potential plaintiffs with greater dispersion in the distribution of their litigation costs will have lower plaintiff win rates.

We implement our model empirically using the reasonable assumption that the distribution of litigation costs for individuals has more variation than the distribution of litigation costs for corporations. Individuals are relatively free to “indulge” their tastes for litigation either by refraining from filing suit where they have a strong case or by filing suit where they do not have a strong case. In contrast, corporations are relatively constrained by market forces and adhere more closely to the goal of profit maximization. Corporations can afford neither to avoid lawsuits that are expected to be profitable nor pursue lawsuits that are not expected to be profitable. Given the greater spread in the distribution of individuals’ litigation costs, we expect that cases where the plaintiff is an individual (relative to cases where the plaintiff is a corporation) to have both a higher trial rate and a lower plaintiff win rate.

Our empirical analysis relies on data gathered from the Administrative Office of the United States Courts that contains information on federal court case filings and outcomes including information on the identity of the parties (individuals or corporations). First we examine variation in trial rates by the identity of the parties. Next we examine how the time it takes to resolve cases by non-trial means (drop or settlement) and the time it takes to reach trial verdicts are related to the identity of the parties. Finally, we examine how plaintiff win rates (both overall and at trial) are related to the identity of the parties.

2. Theoretical Framework

In our view, any plausible model of the litigation process has several properties that are relevant here. First and other things equal, a potential claimant will be more likely to file a lawsuit if the costs of litigation are low. This follows directly from the presumption that only cases with positive expected value to the plaintiff are filed.⁴ Second, conditional on a

⁴ The positive expected value might consist at least in part of nonpecuniary returns from pursuing a claim; e.g., a desire for justice for a perceived wrong.

lawsuit being filed, there will be more trials when the costs of litigation are lower. Third, where litigation costs are lower, potential claimants will be more likely to file claims in which they have a lower probability of prevailing. The intuition for each of these properties is clear: lower costs of disputing will lead to more disputes. We build on these observations in the context of a simple general framework to derive our empirical implications without relying on a specific model of the negotiation process.

Consider the decision of a potential plaintiff regarding whether or not to file a lawsuit. Without being specific about the information structure or timing of the negotiation process, the suit has some expected value to the plaintiff, V_p , as a function of the likelihood that the defendant would be found liable at trial (π), the expected damages that would be awarded at trial conditional on a finding of liability (D), and the costs to the plaintiff and defendant (C_p and C_d respectively) of litigation. The potential plaintiff will file a lawsuit if and only if $V_p \geq 0$.

More formally, the expected value of filing a suit is⁵

$$(1) \quad V_p = V_p(\pi, D, C_p, C_d),$$

The condition for filing then is

$$(2) \quad V_p(\pi, D, C_p, C_d) \geq 0.$$

The key properties of the model of litigation are that V_p is monotonically decreasing in C_p and monotonically increasing in π . The first property implies that the condition for a potential plaintiff to file a lawsuit can be expressed as the plaintiff's litigation costs being less than some threshold value, C_p^* , where C_p^* is a function of π , D , and C_d :

$$(3) \quad C_p < C_p^*(\pi, D, C_d).$$

This simply expresses the reasonable view that the set of lawsuits are selected from the lower tail of the distribution plaintiffs' costs, other things equal. It seems reasonable that

⁵ Depending on the information structure of the game, π , D , and/or C_d may not be known to the plaintiff, *ex ante*, and, if not known, will be represented by the parameters of some prior distribution.

the threshold cost level will be positively related to the likelihood of liability (π) and the stakes of the case (D) so that plaintiffs with lower litigation costs will be more likely to file low π and low D cases. It may be that C_p is increasing in D , and this would at least partially offset the attractiveness of high-stakes cases to potential plaintiffs.

Define the distribution of litigation costs among potential plaintiffs as

$$(4) \quad C_p = \mu + \sigma Z,$$

where Z is a random variable with mean zero and variance one. The mean of litigation costs is μ , and this may depend on the size of the case (π and D). The parameter σ , is a scale parameter which determines the variance of the litigation cost distribution (σ^2). This parameter, which controls the dispersion of plaintiffs' litigation costs in the population of potential claims without affecting the mean, will play a central role in our analysis.

The expected value of litigation costs conditional on a case being filed is

$$(5) \quad \begin{aligned} E(C_p | C_p < C_p^*) &= \mu + \sigma E(Z | \mu + \sigma Z < C_p^*) \\ &= \mu + \sigma E(Z | Z < (C_p^* - \mu)/\sigma) \\ &= \mu + \sigma E(Z | Z < Z^*), \end{aligned}$$

where

$$(6) \quad Z^* = (C_p^* - \mu)/\sigma$$

and is the threshold value expressed in terms of Z . Note that this conditional mean is less than the mean of the unconditional distribution (μ) of costs, so that average litigation costs among suits filed are lower than average litigation costs in the set of all potential claims.

The central characteristic of negotiation models that we rely on is that lower litigation costs among filed cases lead to more trials.⁶ Based on equations 5 and 6, this suggests

⁶ We use the conditional mean ($E(C_p | C_p < C_p^*)$) as an obvious measure of the location of the conditional distribution of costs. It is possible to develop theories of the litigation process where low cost cases are more likely to go to trial but where the mean of the litigation costs distribution would not be a sufficient statistic for these purposes.

that the trial rate is inversely related to the filing threshold, Z^* . But it is difficult to determine how case and plaintiff characteristics affect the filing threshold. The effects of the probability of liability (π) and the value of damages (D) are ambiguous because, while higher probability of winning and higher stakes both increase C_p^* , they also likely increase mean litigation costs (μ).

We can say quite a bit more about the effect of the scale parameter, σ , on the conditional mean of litigation costs. The derivative of the conditional mean of the cost distribution with respect to σ is

$$(7) \quad \frac{\partial E(C_p | C_p < C_p^*)}{\partial \sigma} = E(Z | Z < Z^*) - \frac{\partial E(Z | Z < Z^*)}{\partial Z^*} Z^*$$

Because the unconditional expectation of Z is zero, the first term in this expression ($E(Z | Z < Z^*)$) is negative. The first part of the second term ($\frac{\partial E(Z | Z < Z^*)}{\partial Z^*}$) is positive simply because increasing the right truncation point of a distribution increases the conditional mean. However, the sign of the second part of the second term, Z^* , depends on whether the threshold value of costs (C_p^*) is above or below the mean of the cost distribution (μ).

If the threshold cost value is at or above the mean of the unconditional cost distribution ($C_p^* \geq \mu$), then Z^* is non-negative and an increase in the scale (and, hence, variance) of the litigation cost distribution leads unambiguously to a reduction in the conditional mean of litigation costs. However, given the premise we started with in the introduction, that people are not litigious (so that the rate of filing of lawsuits is generally quite low relative to the pool of claimable incidents) it seems unreasonable that the threshold plaintiff's litigation cost value is greater than the unconditional mean. If the cost distribution were symmetric, then a threshold above the mean would yield a suit rate greater than 0.5. This does not accord with existing evidence.

The more reasonable case is where $C_p^* < \mu$ so that Z^* is negative. Here, if the cost distribution is symmetric, the fact that the threshold is below the mean yields a suit rate less than 0.5. But the effect of scale on the conditional mean is ambiguous in this case. Both terms in equation 7 are negative so that the sign of their difference is indeterminate. Fortunately, there is a reasonable restriction on the distribution of Z that is sufficient for

the needed result (that an increase in σ reduces the conditional mean of litigation costs) and that is satisfied by many common distributions. Rewrite equation 7 by subtracting and adding Z^* . On rearrangement of terms, this yields

$$(8) \quad \frac{\partial E(C_p | C_p < C_p^*)}{\partial \sigma} = [E(Z | Z < Z^*) - Z^*] - \left[\frac{\partial E(Z | Z < Z^*)}{\partial Z^*} - 1 \right] Z^*$$

The first term in brackets is negative by construction. A sufficient condition for the second term in brackets to be negative is that the density function of Z be log-concave.⁷ Examples of distributions satisfying this condition are the normal, uniform, beta, and extreme-value distributions. Since Z^* is negative when $C_p^* < \mu$, log-concavity of the distribution of Z is sufficient for $\frac{\partial E(C_p | C_p < C_p^*)}{\partial \sigma} < 0$.

The conclusion we draw from this discussion is that, under fairly general conditions, an increase in the scale of the unconditional distribution of plaintiffs' litigation costs results in a reduction in average plaintiffs' litigation costs among lawsuits filed. We can operationalize this by considering two groups of potential litigants who have the same mean litigation costs but whose cost distributions have different scale parameters, and, hence, different variances. These two groups seem equally litigious on average, but this is misleading. The process of selection of lawsuits from these two pools of potential litigants yields pools of actual litigants that differ in important ways. Specifically, those litigants drawn from the high-variance distribution will have lower litigation costs, on average, than those litigants drawn from the low-variance distribution.

The central testable implication of this analysis is that plaintiffs drawn from the high-variance distribution will be more likely to push their cases to a trial verdict (higher trial rate). This follows directly from the lower average cost of litigation among plaintiffs drawn from the high-variance distribution.

A second testable implication is that plaintiffs drawn from the high-variance distribution will be less likely to win their lawsuits, either at trial or through a pre-trial settlement.

⁷ A distribution of some random variable X is log concave if the density f satisfies the condition that $f(\lambda x_1 + (1 - \lambda)x_2) \geq [f(x_1)]^\lambda [f(x_2)]^{1-\lambda}$. In other words, the logarithm of the density function must be a concave function. Heckman and Honore (1990) derive the specific result for log-concave distributions used here: $0 \leq \frac{\partial E(Z | Z < Z^*)}{\partial Z^*} \leq 1$.

The heuristic argument for this is that, where average litigation costs are lower (as they are in cases drawn from a pool of potential plaintiffs with a high variance in the cost distribution), lower quality (lower π) cases will meet the criterion for filing ($C_p < C_p^*(\pi, D, C_d)$). More formally, we can write the criterion for case filing in terms of the probability of liability as

$$(9) \quad \pi > \pi^*(C_p, D, C_d).$$

conditional on the value of C_p , which is known to the potential plaintiff. This expression is derived directly from equation 2, noting that V_p is monotonically increasing in π . As long as C_p is unrelated to π , the threshold value of the liability probability, π^* , is positively related to C_p .⁸ Thus, where plaintiffs' litigation costs are lower (as they are for cases filed from a pool of plaintiffs with a high-variance in their cost distribution), the minimum threshold for π is lower and the average quality of cases filed will be lower. The result will be lower plaintiff win rates.

While this prediction on plaintiff win rates refer to the win rate among all lawsuits filed, a natural extension is to plaintiff win rates at trial. While the selection process that yields trials from the pool of lawsuits is surely not random, it seems reasonable that, since plaintiffs with lower costs of litigation are more likely to push their cases to trial, the average quality (from the plaintiff's point of view) of tried cases drawn from the "low-cost" pool of plaintiffs will be lower. The result will be lower plaintiff win rates at trial. We examine both overall plaintiff win rates and plaintiff win rates at trial in our empirical analysis.

Defendants can also be thought of as being drawn from the same two groups as the plaintiffs, but, because defendants do not decide for themselves whether or not they will be involved in litigation, there is only a weaker selection effect of defendants' characteristics on case outcomes.⁹ In fact, the definition of the plaintiff in a lawsuit is the party to the dispute who made the decision to file a claim. It is interesting that most models of the

⁸ Actually, π need not be unrelated to C_p . A weaker condition is that C_p cannot increase too rapidly with decreases in π . The intuition is that weaker cases will be attractive for low-cost litigants to file as long as weaker cases do not cost too much more to litigate.

⁹ There are at least two senses in which defendants do affect the decision regarding litigation involvement.

litigation process start with the set of filed cases and ignore the plaintiff's decision to file. Thus, it is not surprising that the labeling of the parties as plaintiff and defendant in most of these models is arbitrary and of no real substance. Our analysis stands in contrast to this. The potential plaintiff's decision to file a lawsuit has important implications for the characteristics of the parties to lawsuits that affect outcomes in important ways.

3. Proposed Empirical Tests of the Selection Model

We propose tests of the two predictions of our selection model (recapitulated here):

Prediction 1: Trial rates (the fraction of filed cases that are tried to a verdict) will be positively related to the variation in the distribution of plaintiffs' litigation costs in the population of potential claims.

Prediction 2: Plaintiff win rates will be inversely related to the variation in the distribution of plaintiffs' litigation costs in the population of potential claims.

A more general prediction follows from the fact, noted earlier, that, once a lawsuit is filed, the roles that the parties play in the process are symmetric. In contrast, during the initial selection process potential plaintiffs are making decisions regarding suit filing based on their own litigation costs probably without good information about the litigation costs of the potential defendants. Thus, cases filed represent closer to random draws from the distribution of defendants' litigation costs but a strongly selected set of draws from the plaintiff's distribution of litigation costs. Absent the initial selection process, we would expect that the identity of the defendant to have the same effect on case outcomes as the identity of the plaintiff. Evidence to the contrary, that the effect of the identities of the parties on case outcomes differs by role, would suggest that the initial selection process is important. This results in a third prediction:

Prediction 3: The effect on case outcomes of the identity of the plaintiff will differ from the effect on case outcomes of the identity of the defendant only if systematic selection of potential claims for litigation is important.

First, potential defendant's behavior with regard to the issues (e.g., should they comply with the terms of a contract) will affect how likely it is that they are the target of a lawsuit. Second, potential defendant's behavior in pre-litigation discussions and/or negotiations will affect how likely it is that they are the target of a lawsuit (Farber and White, 1994). But neither of these is as direct as the plaintiff's decision regarding case filing.

The key to our empirical tests is the identification of sets of lawsuits that were derived from distributions with different degrees of variation in plaintiff and defendant litigation costs. We argue that whether the parties to a suit are individuals or corporations provides just the kind of separating information we need.

Our working assumption is that corporations face relatively more market discipline than do individuals in their decisions regarding litigation and are not able to deviate substantially from profit-maximizing behavior in their role as potential plaintiffs when evaluating whether or not to file a lawsuit. Thus, the litigation costs of corporations are largely composed of the pecuniary costs associated with litigation: attorney's fees and the value of time associated with undertaking litigation. There is relatively less scope for non-pecuniary factors such as tastes for or against confrontation or litigation to play a role. This means that corporations are relatively likely to file suit where the expected dollar payoff is positive and are relatively likely not to file suit where the expected dollar payoff is negative.

Individuals face a different set of constraints. Individuals are not income or profit maximizers. Individuals are utility maximizers, and it is reasonable to assume that there is substantial variation across individuals with regard to their non-pecuniary tastes for confrontation and litigation. These tastes can be interpreted as contributing both positively or negatively to the costs of litigation. Individuals are free, within limits, to make decisions regarding litigation that take account of these tastes. The relative lack of market discipline on individuals means that variation in tastes for litigation across individuals can contribute strongly to the overall variation in costs of litigation for individuals. In contrast, variation in tastes for litigation across corporations (or their managers) can contribute only in a limited way to the overall variation in costs of litigation for corporations. Thus, the overall variation in costs of litigation is expected to be larger for individuals than for corporations.

We do not want to overstate our claim regarding the constraints faced by corporations relative to individuals. Individuals as potential plaintiffs are free to avoid any litigation they choose, and this can include claims with high expected values. But it may well be the case that individuals are not completely free to pursue low expected value claims. They could well be liquidity constrained in paying for legal representation, and representation is

not likely to be easily obtainable on a contingency basis for low expected value claims.¹⁰ In addition, not all corporations are run at arms length strictly for the pecuniary benefit of shareholders. There are many closely-held corporations that are managed by their owners and whose resources can be used to maximize the utility (as opposed to dollar value) of their owners. Thus, some corporations will behave more like individuals in their dealings with the legal system. The force of our assumption is that *on average* individuals are more likely than corporations to consider non-pecuniary factors when making decisions regarding litigation, and this shows up as more variation in the distribution of litigation costs for individuals than in the distribution of litigation costs for corporations.

It is important to emphasize that individuals' ability to exercise their tastes regarding litigation does not necessarily mean that individuals are more litigious than corporations. As noted above, there is evidence that most individuals with valid claims do not pursue these claims (Danzon, 1985; Report of the Harvard Medical Practice Study, 1990; Hensler et al., 1991). This is consistent with most individuals having tastes regarding litigation that contribute positively to total litigation costs. Our primary theoretical concern and our empirical implications stem from differences in the variation in costs and not from differences in the level of costs.

We can restate our two specific testable predictions in terms of the identity of the plaintiff:

Prediction 1: Trial rates will be higher for cases where the plaintiff is an individual than in cases where the plaintiff is a corporation.

Prediction 2: Plaintiff win rates will be lower for cases where the plaintiff is an individual. The third prediction, that it is only (or, more modestly, largely) through selection of cases for litigation that case outcomes will be affected differently by the identity of the plaintiff than by the identity of the defendant, does not require the assumption of a specific difference between individuals and corporations.

¹⁰ One result of this could be lower quality representation for low expected value claims. We discuss this below when considering alternative interpretations of our results.

4. The Data and Summary Statistics

To implement our empirical tests, we use data gathered by the Administrative Office of the United States Courts. When a civil case terminates in federal district court, the court clerk files a form with the Administrative Office containing information about the case (Admin. Office of the U.S. Courts, 1985).¹¹ The form includes data regarding the subject matter and jurisdictional basis of the case, the dates of filing and termination, the procedural progress of the case at termination, including whether it was tried, and, where a judgment was entered, who prevailed.

Federal jurisdiction for many cases filed in federal court is not based on federal law. Plaintiffs with a state law course of action involving more than \$50,000 may file suit in federal court if the plaintiff and the defendant are from different states.¹² These cases are said to be diversity cases because federal jurisdiction is based on the diversity of citizenship between the parties. In federal court cases based on diversity jurisdiction, the Administrative Office has, since 1986, tracked whether the plaintiffs and defendants are individuals or corporations.¹³ This information is not collected in cases other than diversity cases. Thus, because this information is crucial to our analysis, we restrict our analysis to diversity cases.

The available data shape the categories of cases in our sample. Since all of the cases in the sample are diversity cases, large classes of cases do not appear in the sample. For example, there are no civil rights, patent, or antitrust cases. These cases may be brought in federal court because they are based on federal law, regardless of the amount in controversy.

The case categories comprising diversity litigation are the common law areas of tort, contract, and property. Our sample contains all federal diversity cases filed between July 1, 1986 and September 30, 1994. We delete all personal injury cases from our sample

¹¹ A complete description of the Administrative Office data appears in Inter-university Consortium for Political and Social Research [ICPSR], 1993

¹² 28 United States Code section 1332.

¹³ The coding scheme identifies the status of the "principal parties" to the lawsuit. No effort is made to have multiple designation of parties to cases, even when cases involve multiple parties (ICPSR, 1993). Strengths and weaknesses of the data are discussed in Eisenberg & Schwab, 1987; Schwab & Eisenberg, 1988.

because only a very small fraction of personal injury cases have corporate plaintiffs. For this reason, it is our view that the experiment of comparing case outcomes by identity of plaintiff makes little sense in personal injury cases.¹⁴ We also exclude a small number of RICO cases and a minor class of diversity cases classified as "other". The remaining sample includes 218,120 cases.

Since our data end with the close of the 1994 federal government fiscal year, we have no data on case progress after September 30, 1994. Thus, any cases not resolved by that date are censored in the sense that their outcomes are not observed. Of the 218,120 cases filed between July 1986 and September 1994, 196,441 were resolved and 21,679 were not resolved in time for the resolution to be recorded in our data. This is an important statistical problem in the analysis of trial rates and outcomes because cases that proceed to trial are likely to take substantially longer to resolve than cases that are dropped or settled without a trial verdict. Thus, tried cases will be censored disproportionately.

Table 1 shows the number of cases in our sample filed in each calendar year.¹⁵ There has been some drop-off in the number of diversity cases since 1986. Table 1 also shows the fraction of cases pending as of September 30, 1994 by year of filing. Overall, about 10 percent of cases in our sample are censored, and it is clear that cases filed more recently are much more likely to be censored.

That this censoring is an important problem is apparent from the tabulation of the fraction of cases tried by year of filing contained in the last column of table 1. This fraction is computed as the fraction of resolved cases where there was a trial verdict. The trial rate computed this way falls dramatically with the year of filing. It is doubtful that this represents a real decline in trial rates. It is almost surely an artifact of the decline in time until the censoring date with year of filing and the concomitant lack of time for cases that will ultimately be tried to reach a trial verdict.

¹⁴ By itself, lack of variation in the identity of the plaintiff does not require the elimination of these cases. But fully 96 percent of personal injury cases have individual plaintiffs compared with 51 percent of other types of cases. No category of cases included in our study has even 80 percent of cases derived from one category of plaintiff.

¹⁵ The number of cases filed in 1986 is about half of the number in the immediately succeeding years because we have no data on cases filed prior to July 1, 1986. Similarly, the number of cases filed in 1994 is about 3/4 the number of cases filed in the immediately preceding years because we have no data on cases filed after September 30, 1994.

One approach to the censoring problem would be to use only cases filed at sufficient temporal remove from the censoring date so that the censoring problem would be minimal. However, the overall trial rate is only 3.58 percent (though this is a somewhat downward biased estimate given the censoring). Even a small amount of censoring is likely to seriously affect estimates of the trial rate. Thus, the fact that 4.5 percent of cases filed in 1991 are unresolved make even this year unsuitable for the analysis of trial rates. One would have to limit the analysis to 1986-1990 to be reasonably sure that censoring would not be a significant problem. While the simple tabulations presented in this section use data for the entire 1986-94 time period and do not account for the censoring, the multivariate statistical analyses presented in succeeding sections address this problem in several ways that we describe below.

While measuring the trial rate is straightforward, determining or even defining who wins is not. The data are fairly limited on this question, and we define a party as winning if a judgement was entered for that party. There are at least three problems with this measure. First, just because a judgement was entered for a particular party does not mean that the party, in fact, "won". For example, a plaintiff who files suit expecting a substantial damage award may settle for or receive a very small damage award. This case would be recored as a judgement for the plaintiff, and we call this a plaintiff win. But the plaintiff might perceive this as a loss. Second, the party for whom a judgement is entered is not coded in the majority (74.4 percent) of the cases resolved without a trial verdict. These are largely cases that are dropped or settled without having a formal judgment entered in court. We proceed ignoring this missing data problem, and the implicit assumption is that the cases for which judgement is not available are not systematically different in the relevant dimensions from those for which data on judgement is available. This is less of a problem for cases resolved with a trial verdict, where only 9.4 percent of judgements are missing. Finally, judgement is sometimes entered for both parties (in about 5 percent of the cases for which we have data). It is not clear who won in these cases, and we proceed as if these are missing as well.¹⁶

¹⁶ We return to the issue of judgement for both parties when presenting the multivariate analysis of win rates.

The data record the broad legal category of each case. Table 2 lists these categories and shows 1) the number of cases in each category in our sample, 2) the number of uncensored cases in each category, 3) the rate at which uncensored cases in each category are resolved at trial, 4) the rate at which uncensored cases in each category are won by the plaintiffs, and 5) the rate at which uncensored cases ending in trial judgments are won by plaintiffs. By far the largest category of cases is “other contract actions”. Additionally, insurance, foreclosure, and negotiable instruments are large case categories. There is a substantial difference in raw trial rates across categories with the highest trial rates in product liability torts and property damage, product liability and the lowest trial rates in foreclosure and negotiable instruments. Plaintiff win rates vary substantially across case categories, from a low of 28 percent in product liability torts to a high of 99 percent in foreclosure cases. Plaintiff win rates at trial differ less across categories, but there are no extreme values. The highest plaintiff win rate is 69.4 percent in negotiable instruments, and the lowest plaintiff win rate is 43.8 percent in “property damage - product liability”. The smaller spread in win rates at trial relative to win rates among all lawsuits filed reflects the systematic process of selection of cases for trial.

Table 3 presents preliminary evidence based on simple cross-tabulations of how the identity of the plaintiff is related to case outcomes. While this analysis does not account for the censoring problem, the results are clear and consistent with our expectations. The trial rate is significantly higher among cases with individual plaintiffs than among cases with corporate plaintiffs. The trial rates for the two groups are 0.0412 and 0.0304 respectively, and the p-value of the difference is < 0.00001 . The overall plaintiff win rate is substantially lower among cases with individual plaintiffs than among cases with corporate plaintiffs (0.608 vs. 0.836, p-value < 0.00001). The plaintiff win rate at trial is also significantly (p-value < 0.00001) lower for cases with individual plaintiffs, but the difference is much smaller (0.586 vs. 0.648).

Table 3 also presents preliminary evidence on how the identity of the plaintiff is related to trial outcomes controlling for the identity of the defendant. These tabulations are again consistent with both predictions of the theoretical model: 1) Controlling for the identity of the defendant, trial rates are higher where the plaintiff is an individual (p-value $< .000001$)

and 2) controlling for the identity of the defendant, plaintiff win rates at trial are lower where the plaintiff is an individual (p-value < .000001).¹⁷

Overall, the specific evidence on how case outcomes vary with the identity of the plaintiff is consistent with the case selection framework we developed coupled with the assumption that individuals are more variable in their litigation costs than are corporations.

Tabulations of case outcomes broken down by the identity of the defendant are also presented in table 3. There are significant differences here as well, with lower trial rates and higher plaintiff win rates among cases with individual defendants. Our theoretical framework was silent on how these outcomes might be related to the identity of the defendant, but it is interesting that individual litigants seem to fare worse (have lower win rates) than do corporate litigants, regardless of their particular role. When corporations sue individuals, the plaintiff wins over 90 percent of the time. In contrast, when individuals sue corporations, the plaintiff wins only 50 percent of the time. When the identities are the same (individual-individual or corporation-corporation) the plaintiff wins about 75 percent of the time. More generally, the tabulations suggest that differences in outcomes by identity of the defendant are of roughly the same magnitude as those by identity of the plaintiff. Thus, the general test for plaintiff selection of cases for litigation, outlined in the previous section (Prediction 3), shows no evidence of this selection.

Finally, it is interesting that there is an inverse monotonic relationship across the four plaintiff/defendant types between trial rates and plaintiff win rates: high trial rates are associated with low plaintiff win rates.¹⁸ This is a natural consequence of systematic selection of potential claims for litigation on the basis of plaintiffs costs. Lower plaintiff costs imply higher trial rates and lower plaintiff win rates.

¹⁷ The p-values reported here are derived from Pearson χ^2 statistics from pairs of two-by-two tables for each hypothesis. For example, for the first prediction, there is a two-by-two table of plaintiff identity (plaintiff or individual) by trial (trial or no trial) for each of the two types of defendants (plaintiff or individual). The test statistic is derived by summing the χ^2 statistic for each table, and the resulting statistic is distributed as χ^2 with two degrees of freedom. All individual two-by-two tables deviate significantly from independence at conventional levels with the exception of the plaintiff win rate by plaintiff type for individual defendants (p-value = 0.177).

¹⁸ Note that this requires no specific assumption about how the litigation costs of individuals differ from those of corporations.

5. Trial Rates – Probit Analysis

In this section we present estimates of probit models of the probability of trial controlling for case category. We address the censoring problem from several perspectives. First, we augment the probit model estimated over the set of resolved (uncensored) cases with a set of dummy variables for filing year in order to account for the fact that cases that go to trial take longer so that resolved cases filed later (closer to the censoring date) will appear to have lower trial rates. Second, we estimate the probit model using only data for resolved cases filed in 1990 or earlier. Only a small fraction of cases this old are censored (see table 1). Finally, we present estimates of a two-equation probit model estimated using data on all cases, censored or not, of the joint probability of trial and censoring in order to derive consistent estimates of the effect of plaintiff and defendant type on the probability of trial.

The basic vector of variables in both equations includes a constant and a set of three dummy variables for party identity (plaintiff individual, defendant individual, both individual). Thus, the base category consists of cases where both parties are corporations, and the key parameter of interest is the dummy variable for cases where the plaintiff is an individual. We also estimate a “full” specification where we add fourteen dummy variables for the fifteen case categories to each vector in order to control for differences across categories in trial rates.

The first two columns of table 4 contain estimates of the two specifications for a simple probit model of the probability of trial using data on the 196441 cases filed between 1986 and 1994 that were resolved by September 30, 1994. In order to account crudely for the censoring at this date, these specifications include a set of dummy variables for year of filing. Where case category is not controlled for (column 1), the estimates suggest that cases with individual plaintiffs are significantly more likely to go to trial than are cases with corporate plaintiffs. Given an average probability of trial of about 0.05, the probit coefficients multiplied by 0.5 can be interpreted approximately as the proportional effect of the relevant variable on the probability of trial. Thus, the estimates suggest that individual plaintiffs are about 6.5 percent more likely to take cases to trial than are corporate plaintiffs. Seemingly offsetting this, the estimates in column 1 suggest that individual defendants are

about 6.5 percent less likely to go to trial than are corporate defendants. Given the finding that the coefficients on the plaintiff-individual and defendant-individual dummies sum to zero and the finding that the coefficient on the 'both-individual dummy is not significantly different from zero, cases where both parties are individuals and cases where both parties are corporations are equally likely to go to trial. These results are consistent with the simple tabulations in table 3.

The picture changes somewhat when case category is controlled for in column 2.¹⁹ Our central prediction, that cases where plaintiffs are individuals are more likely to go trial, continues to be supported by the evidence, though the marginal effect on the probability falls to five percent. However, the identity of the defendant no longer has a significant effect on the trial rate. This pattern of results, that the identity of the plaintiff matters while that of the defendant does not, is consistent with the general prediction of the selection model, that the effect of the identity of the plaintiff on case outcomes will differ from the effect of the identity of the defendant.

Given the substantial amount of censoring in the data, we next reestimated the model using data only on the 132086 cases filed in 1990 or earlier and resolved by September 30, 1994. The censoring problem is much smaller in these data because these cases have all had at least four years to be resolved. Table 1 contains censoring rates by year of filing. A summary is that the censoring rate for cases filed in 1990 or earlier is 0.46 percent compared with 24.7 percent for cases filed after 1990. Estimates of the two probit specifications for the restricted sample are contained in columns 3 and 4 of table 4. The results are virtually identical to those for the full sample. This suggests that, while there is obviously much censoring in the full sample, it does not appear to bias the results.

Finally, we estimated a structural bivariate probit model that jointly determines the trial and censoring probabilities. A key to identification of the trial equation in the face of the censoring is to find variables that affect the censoring process but are not related to

¹⁹ The improvement in the likelihood function when case categories are controlled for is dramatic and significant. This is not surprising given the sharp differences in trial rates across case categories that we found in table 2.

the probability of trial.²⁰ A natural variable that fits this description is the potential time available for case resolution. This is the number of days between the date the suit was filed and the censoring date (September 30, 1994). In the analysis that follows we include an unrestricted cubic function of the potential time in the vector of variables in the censoring equation.²¹

Columns 5 and 6 of table 4 contains estimates of the key parameters of the model for the two different specifications.²² We present only the estimates of the trial equation, but the estimates of the censoring equation show clearly that potential time is a powerful determinant of the probability of censoring, with longer potential time yielding substantially reduced censoring probabilities. As before, the case category dummies are significant determinants of outcomes, with the log-likelihood improving dramatically when the case category dummy variables are included. Not surprisingly, the estimated value of the correlation between the errors in the trial and censoring equation is strongly positive, suggesting that cases that are more likely to go to trial for unobservable reasons are also more likely to be censored. Despite this endogenous censoring process, the estimates from the bivariate probit model with regard to trial rates and party identity are virtually identical to those from the univariate probit models: cases with individual plaintiffs are about 9 percent more likely to go to trial and the identity of the defendant does not have a significant effect on the trial rate.

Overall, the evidence on trial rates is clear and show evidence consistent with the selection process we outlined. Cases with individual plaintiffs are more likely to go to trial. In combination with this, our finding that the identity of the defendant is not related to trial rates is consistent with our general test of the selection process: Controlling for case

²⁰ Such exclusions are not strictly required, but, without such exclusions, identification relies solely on nonlinearities in the probability functions associated with the joint distribution of the errors.

²¹ We also experimented with a set of dummy variables for the year of case filing. This did not yield nearly as good a fit as the cubic function (using a log-likelihood criterion). The results with regard to the party-identity variables were not affected by the particular measure of potential time used.

²² These estimates were obtained using fortran programs implementing the algorithm described by Berndt, Hall, Hall, and Hausman (1974).

category, the identity of the plaintiff and the defendant have different effects on the trial rate

6. Time to Resolution: A Competing Risk Approach

Our finding, that cases with individual plaintiffs are more likely to go to trial, raises some interesting questions regarding how the time to reach a settlement and the time to reach a trial verdict are related to party identity.²³ While our theory is silent with regard to actual time, this is an important question given the long processing time of cases implicit in the censoring rates in table 1.²⁴ We investigate this in the context of a competing risk model where settlement and trial are competing “risks” for case resolution.

We choose a particularly simple specification where the two durations are assumed to be jointly log-normal. We specify

$$(10) \quad \ln D_{si} = X_{1i}\beta_1 + \epsilon_{1i},$$

and

$$(11) \quad \ln D_{ti} = X_{2i}\beta_2 + \epsilon_{2i},$$

where $\ln D_{si}$ and $\ln D_{ti}$ are the log times to settlement and trial respectively, X is a vector of case/claimant characteristics and β_1 and β_2 are parameter vectors. We assume that the error terms (the ϵ 's) have a bivariate normal distribution.

In the competing risk model it is important to keep track of what is known about the two durations in each of the three configurations of the data (censored, settled, tried). Where the case outcome is censored, we know that both durations were longer than the potential time (the time between the filing data and September 30, 1994). Denote this potential time as D_{pi} . Thus, we know that $\ln D_{si} > D_{pi}$ and $\ln D_{ti} > D_{pi}$. Where there is

²³ The settlement time refers to the time from case filing to resolution without at trial verdict. Many “settled” cases are dropped without any payment to the plaintiff or are dismissed by the court. The trial time is the time from from case filing to resolution through a trial verdict.

²⁴ Some simple statistics on mean time to resolution for cases filed in 1990 or earlier (in order to minimize the censoring problem) are that mean time to a settlement is 328 days (s.e. = 0.688) while mean time to a trial verdict is 643 days (s.e. = 3.57). The median times to resolution are 207 and 509 days respectively.

is a settlement at D_i , we know that $\ln D_{s,i} = D_i$ and that $\ln D_{t,i} > D_i$. Finally, where there is a trial judgement at D_i , we know that $\ln D_{t,i} = D_i$ and that $\ln D_{s,i} > D_i$. These relationships are used in conjunction with the assumption of joint normality of the errors in equations 10 and 11 to derive the appropriate likelihood function. An interesting feature of the competing risk approach is that it naturally takes account of the potential time available to resolve a case in computing the probability of censoring.

Table 5 contains maximum-likelihood estimates of the key parameters of the competing risk model for the two specifications of the X vectors used in the probit models. As before, the first, in columns (1) and (2), includes only the basic variables, and the second, in columns (3) and (4) is the full model including dummy variables for case category as well. Not surprisingly, the controls for case category account for a significant fraction of the variation in the model (p-value from a likelihood ratio test < 0.00001). Thus, we focus the discussion on the results in columns (3) and (4) for the model including the case category controls.

The settlement time is not significantly related to the identity of the plaintiff. However, settlement time is more strongly related to the identity of the defendant. Settlement time is about 9 percent shorter where the defendant is an individual.²⁵ The time until a trial verdict is strongly related to the identity of the plaintiff. This duration is about 30 percent lower where the plaintiff is an individual. Time until a trial verdict is insignificantly positively related to the identity of the defendant. These results are fully consistent with the probit results. The 30 percent shorter time until a trial verdict yields the predicted higher trial rate in cases with individual plaintiffs.

The estimates also show that cases individual defendants settle about 9 percent more quickly. Given that times to trial verdict are so much longer than times until settlement, this relatively small difference does not yield significant differences in trial rates by the identity of the defendant.

Finally, note that unobservable factors that affect time to settlement and trial verdict (the ϵ 's in equations 10 and 11) are strongly negatively correlated ($\hat{\rho} = -0.89$). One

²⁵ Given that the dependent variable is log duration, the coefficient on a dummy variables can be interpreted approximately as the proportional effect (as long as the coefficient is less than about 0.25 in absolute value).

plausible interpretation of this result is that cases that are not likely to settle (have long settlement durations) go to trial relatively quickly. Similarly cases that are not likely to go to trial (have long times to a trial verdict) settle relatively quickly.

The central insight gained from these estimates is that the source of the higher trial rate in cases with individual plaintiffs is that trials happen much more quickly and not that settlement is delayed.

7. Plaintiff Win Rates

Finally, we analyze plaintiff win rates. We analyze data on who wins at trial only for the set of cases for which information the party for whom a judgement was entered was recorded. We noted above that there is a substantial amount of missing data, and we attempt no statistical correction for the selection process that might yields the set of cases with observed judgements.²⁶ On this basis, we present simple probit analyses of the probability that a judgement is entered for the plaintiff for the set of cases that were tried to a verdict and for which a winner was recorded. As before, we treat the censoring problem in this analysis informally. We take two approaches. First, we include a set of dummy variables for year of filing. Second, we reestimate the model over the subset of suits filed no later than 1990 (where the censoring problem is minimal).

Judgements were entered for both parties in a small fraction (5 percent) of the cases. These observations were not included in our analysis of plaintiff win rates because they represent a clear win for neither party. We did investigate this further using an ordered probit model where judgement for both parties was considered intermediate between a judgement for the defendant and a judgement for the plaintiff. These estimates are not presented here, but there are substantively identical to the simple probit estimates of the probability of a judgement for the plaintiff that we do present.

²⁶ We do not present results of some standard techniques for potential bias introduced by the missing data because we lack a convincing instrument that is correlated with the probability that data are missing but is not correlated with who wins at trial. Such an analysis would rely entirely on the functional form of the probability functions or on unsupported exclusion restrictions. Simple tabulations show that the missing data rate does not differ substantially with the identity of the plaintiff and defendant with one exception: judgements are less likely to be missing for cases where corporations sue individuals than in the other three categories (60 percent vs. 77 percent).

Table 6 contains estimates of the probit models for the probability that a judgement is entered for the plaintiff. The estimates in column 1 are for the basic specification of the model. The estimates in column 2 are for a model that includes case category controls as well. Since the model that includes the additional controls fits significantly better (p-value < 0.00001) than the model without these controls, we focus on the estimates in column 2. These estimates show that individual plaintiffs are significantly less likely to win a judgement (about 25 percentage points less likely at the mean) than are corporate plaintiffs (p-value < 0.0001). The estimates also show that individual defendants are significantly less likely to win a judgement (about 15 percent less likely at the mean) than are corporate defendants (p-value < 0.0001).

While we do not expect that censoring is a direct problem in estimating the determinants of who wins, we also reestimated the model using only cases filed in 1990 or earlier years, as before. These estimates are contained in columns 3 and 4 of table 6. Focusing on the results in column 4 that control for case category, we find results that are virtually identical to those we found using all years.

These results are consistent with the specific prediction of the theoretical framework that the selection process generates a set cases where individual plaintiffs have lower win rates than do corporate plaintiffs. It is striking that we also find that individual defendants have lower win rates than do corporate defendants. Our specific theory of the case selection process was silent on this issue, but the general test of selection that we outlined suggested that, to the extent that selection of cases for litigation is important, the identity of the plaintiff would have a different effect on win rates than the identity of the defendant. Our evidence is mixed on this point. The coefficients on the individual-plaintiff variable and the individual-defendant variable are significantly different in absolute value from each other (p-value < 0.0001), but the qualitative nature of the results suggests that, regardless of the role they play, individuals fare worse in litigation than do corporations,

Finally, we repeated the analysis of the probability of plaintiff wins using only the cases that were decided by a trial verdict. These estimates are contained in columns 5-8 of table 6. This analysis of win rates at trial shows smaller and relatively imprecisely estimated effects of party identity on the probability of plaintiff win. This is not surprising given

the small number of cases decided at trial. But the qualitative nature of the results is similar to that found with the large sample. Using the estimates in column 6 (all years, controlling for case case categories), individuals are less likely (about 4 percent) to prevail at trial whether they are plaintiffs or defendants, but the effect does not differ significantly in absolute value by role. The estimates for the cases filed in 1986-90 (column 8) based on an even smaller sample, do not show a statistically significant effect of the identity of the plaintiff on the plaintiff win rate at trial. However, individual defendants remain significantly less likely to prevail at trial.

In summary, our analysis of plaintiff win rates, both overall and at trial, is consistent with the specific implications of our case selection framework. Individual plaintiffs are less likely to win a judgement than are corporate plaintiffs. With regard to the general test of the case selection model, based on whether the identity of the plaintiff has different effects on win rates than the identity of the defendant, while individuals are less likely to win a judgement regardless of the role they play, we do find that the effect is somewhat smaller for defendants than for plaintiffs.

8. Alternative Explanations

While we highlight variation in the distribution of litigiousness in our analysis of trial rates and trial outcomes, we recognize that other factors can be as important or even more important in explaining litigation patterns. For example, several other factors also predict differences in plaintiff/defendant success rates at trial. Such factors as asymmetric stakes to plaintiffs and defendants (Waldfogel, 1995), the have and have-not status of plaintiffs and defendants (Galanter, 1974), differing information available to plaintiffs and defendants (Daugherty & Reinganum, 1993; Hay, 1995; Hylton, 1993; Spier, 1992), agency problems (Miller, 1987), and differing quality of legal representation (Schwab & Eisenberg, 1988) are possible bases for explaining the pattern of plaintiff trial success rates.

Our model's predictions about both trial rates and success rates in litigation as they relate to the identity of both plaintiffs and defendants helps to separate its implications from factors emphasized in other models. Some factors generate no clear prediction about trial rates. For example, consider the possibility that individual plaintiffs' low success

rates in cases pursued to judgment are a consequence of the fact that individuals obtain weaker counsel than corporations. But this theory supplies no clear guidance with respect to trial rates. If weak lawyers push individual plaintiffs to a trial judgement it might be reasonable to expect weak lawyers to push individual defendants to a trial judgement. But our evidence (table 4) suggests that cases with individual defendants are less likely to be tried. On the other hand, if the lower win rate of individuals is due to lower quality counsel, individual defendants also should have lower win rates at trial. Our evidence is consistent with this.

The fact that litigation rates show significant correlation largely with individual plaintiff status, and not with individual defendant status, suggests that something more than weaker lawyering for individuals is going on. Differing litigiousness is a possible explanation, but the fact that plaintiffs make the key initial decision regarding whether or not to file suit suggests that the theory of selection we present in this study is a sensible overall explanation.

To the extent that individual plaintiffs are relatively more likely to have attorneys representing them who are paid on a contingency basis as opposed to an hourly basis, there might be an agency explanation for differences in outcomes (Miller, 1987).²⁷ Since trials require disproportionately more resources than the pre-trial phase, attorneys being compensated on contingency would be more likely to recommend settlement while attorneys being paid by the hour would be more likely to push cases to trial. This would suggest that individual plaintiffs, being disproportionately represented by attorneys paid on a contingency basis, would have *lower* trial rates than corporate plaintiffs. This is precisely the opposite of what our selection theory predicts and what we found in the data.

The have and have-not explanation also forecasts either different litigation rates than those observed or provides only ambiguous predictions. If individual plaintiffs are have-nots relative to corporations, their trial rates may be expected to be lower, not higher, than corporate plaintiffs. This is because the have-nots, by definition, have more limited

²⁷ Note that we have omitted personal injury cases from our analysis and that this is the predominate area where contingency compensation for plaintiffs' attorneys is used (Hensler et al., 1991: 136; Kritzer, 1990: 59). But contingency fees are also common for individuals' contract cases. Id.

resources than corporations and may be forced to drop their cases before trial. On the other hand, corporate defendants, knowing of the limited resources of their individual plaintiff adversaries, may press cases to trial precisely to drain plaintiffs of their limited resources. Without a model, it is difficult to sort out the net effect of these contrasting tendencies on trial rates.

9. Summary and Concluding Remarks

Empirical analysis of the selection of cases for litigation is difficult because only a small fraction of the vast set of potential claims result in the filing of a lawsuit. And those potential claims that do not result in lawsuits not only are not observed, they are not even enumerated. Yet the process through which cases are selected for litigation cannot be ignored because it yields a set of lawsuits and plaintiffs that is far from a random selection either of potential claims or, perhaps more to the point, of potential claimants. In this study, we developed a theoretical framework that provides the conceptual “lever” needed to analyze the selection process empirically despite the absence of information on potential claims not filed. We use this framework to investigate how the selection process affects the characteristics of the claimants in suits filed in relation to the underlying distribution of potential claimants. We also use this framework further investigate the progress of these suits through the litigation process.

We implemented the model empirically by assuming that individuals vary more in their litigiousness (inverse costs of litigation) than do corporations. This assumption, coupled with the case selection process we derived theoretically, yielded clear predictions on trial rates and win rates as a function of whether the plaintiff and defendant were individuals or corporations. Our empirical analysis, using data on over 200,000 federal civil litigations, yields results that are strongly consistent with the theory. Lawsuits where the plaintiff is an individual are found to have higher trial rates than lawsuits where the plaintiff is a corporation. Consistent with this, there is also evidence that individual plaintiffs file weaker cases than corporate plaintiffs in that individual plaintiffs win rates are lower.

More generally, the evidence strongly suggests that trial rates are substantially affected by the identity of the plaintiff but not by the identity of the defendant. This is evidence that

the selection of claims for litigation, which is made by potential plaintiffs, is an important phenomenon.

Overall, our empirical evidence is consistent with the view that potential claims are selected for litigation based on the litigiousness of the potential plaintiff and that this selection process has important implications for patterns of case outcomes. These findings should open new areas of exploration. For example, large differences in claiming rates, such as the difference in claiming rates between automobile accident cases and other tort cases (Hensler et al., 1991), might be based on differences in the distribution of litigation costs. To the extent that this is the case, our conception of the case selection process could prove very useful in understanding differences in the pattern of case outcomes. For example, there is substantial variation in plaintiffs' trial success rate across tort categories (Clermont & Eisenberg, 1992). Some of this variation may be linked to unobserved differences in litigation cost distributions that have an effect through the case selection process. Linking our findings about the influence of the selection of cases for filing to outcomes in other classes of cases seems like fertile ground for future research.

10. References

- Administrative Office of the U.S. Courts, *Guide to Judiciary Policies and Procedures transmittal 64*, vol. XI, at II-19 to -28 (Mar. 1, 1985).
- Berndt, E.K.; Hall, B.H.; Hall, R.E.; and Hausman, J.A., "Estimation and Inference in Nonlinear Structural Models," *3/4 Annals of Economic and Social Measurement* 653 (1974).
- Clermont, Kevin M. & Theodore Eisenberg, "Trial by Jury or Judge: Transcending Empiricism," *77 Cornell L. Rev.* 1124 (1992).
- Cooter, Robert D. & Daniel L. Rubinfeld, "Economic Analysis of Legal Disputes and Their Resolution," *27 J. Econ. Lit.* 1067 (1989).
- Danzon, Patricia M., *Medical Malpractice: Theory, Evidence and Public Policy* (1985). Harvard Univ. Press, Cambridge, MA.
- Daugherty, Andrew F. & Jennifer F. Reinganum, "Endogenous Sequencing in Models of Settlement and Litigation," *9 J. L. Econ. & Org.* 314 (1993).
- Eisenberg, Theodore, "Testing the Selection Effect: A New Theoretical Framework with Empirical Tests," *19 J. Leg. Studies* 337 (1990).
- Eisenberg, Theodore & Stewart Schwab, "The Reality of Constitutional Tort Litigation." *72 Cornell L. Rev.* 641 (1987).
- Farber, Henry S. & Harry C. Katz, "Interest Arbitration, Outcomes, and the Incentive to Bargain," *33 Industrial and Labor Relations Review* 55 (1979).
- Farber, Henry S. & Michelle J. White, "Medical Malpractice: An Empirical Examination of the Litigation Process," *22 Rand Journal of Economics* 199 (1991).
- Farber, Henry S. & Michelle J. White, "A Comparison of Formal and Informal Dispute Resolution in Medical Malpractice," *23 J. Leg. Studies* 554 (1994).
- Hay, Bruce L., "Effort, Information, Settlement, Trial," *24 J. Leg. Studies* 29 (1995).
- Heckman, James J. and Bo E. Honore, "The Empirical Content of the Roy Model." *58 Econometrica* 1121 (1990).

- Hensler, Deborah R., M. Susan Marquis, Allan F. Abrahamse, Sandra H. Berry, Patricia A. Ebener, Elizabeth G. Lewis, E. Allan Lind, Robert J. MacCoun, Willard G. Manning, Jeanette A. Rogowski, Mary E. Vaiana, *Compensation for Accidental Injuries in the United States* (1991). RAND Institute for Civil Justice.
- The Report of the Harvard Medical Practice Study, Patients, Doctors, and Lawyers: Medical Injury, Malpractice Litigation, and Patient Compensation in New York* (1990). President and Fellows of Harvard College.
- Galanter, Marc. "Why the Haves" Come Out Ahead: Speculations on the Limits of Legal Change," 9 *Law and Soc'y Rev.* 95 (1974).
- Hylton, Keith N., "Asymmetric Information and the Selection of Disputes for Litigation." 22 *J. Leg. Studies* 187 (1993).
- Inter-university Consortium for Political and Social Research [ICPSR], (Federal Judicial Center, Principal Investigator), *Federal Court Cases: Integrated Data Base, 1970-1987*. ICPSR 8429 (5th ICPSR ed. March 1993).
- Kritzer, Herbert M., *The Justice Broker: Lawyers and Ordinary Litigation* (1990). Oxford University Press, New York.
- Miller, Geoffrey P., "Introduction: Economic Analysis of Civil Procedure," 23 *J. Leg. Studies* 303 (1994).
- Miller, Geoffrey P., "Some Agency Problems in Settlement," 16 *J. Leg. Studies* 189 (1987).
- Priest, George L. & Benjamin Klein, "The Selection of Disputes for Litigation." 13 *J. Leg. Studies* 1 (1984)
- Shavell, Steven, Suit, "Settlement and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs," 11 *J. Leg. Studies* 55 (1982).
- Schwab, Stewart J. & Theodore Eisenberg, "Explaining Constitutional Tort Litigation: The Influence of the Attorney Fees Statute and the Government as Defendant." 73 *Cornell L. Rev.* 719 (1988).
- Spier, Kathryn E., "The Dynamics of Pretrial Negotiation," 59 *Rev. Econ. Stud.* 93 (1992).
- Waldfogel, Joel, "The Selection Hypothesis and the Relationship between Trial and Plaintiff Victory," 103 *Journal of Political Economy*, 229 (1995).

TABLE 1
Case Distribution and Resolution by Year of Filing

Year Filed	Number of Cases Filed	Fraction Pending (9/30/94)	Fraction Tried
1986	14569	.0011	.0454
1987	31484	.0009	.0424
1988	33903	.0024	.0373
1989	28918	.0051	.0359
1990	23826	.0144	.0376
1991	24310	.0450	.0391
1992	22789	.1003	.0301
1993	21960	.2772	.0198
1994	16361	.7089	.0065
Total	218120	.0994	.0358

Note: *Fraction Pending* is computed as the fraction of *Number of Cases Filed* that are not resolved by September 30, 1994 *Fraction Tried* is computed as the fraction of cases resolved by September 30, 1994 that were tried to a verdict.

TABLE 2
Case Distribution and Outcomes by Case Category

Case Category	Number of Cases	Number Resolved	Trial Rate (Resolved Cases)	Plntf Win Rate (All)	Plntf Win Rate (Tried)
Insurance	47640	41777	.0465	.481	.536
Marine	1064	938	.0448	.759	.486
Neg Instruments	10260	9562	.0235	.909	.665
Recov of Overpay	847	787	.0343	.816	.682
Stockholder Suits	1241	1098	.0419	.472	.649
Other Contract	120169	108769	.0340	.754	.670
Contract Prod Liab	2426	2173	.0437	.541	.578
Foreclosure	13329	12975	.0019	.990	.667
Rent, Lease, Eject	1348	1254	.0367	.772	.683
Torts to Land	1630	1371	.0569	.395	.424
Tort, Product Liab.	442	365	.121	.281	.361
Other Real Prop	2798	2523	.0547	.429	.566
Other Fraud	6913	5953	.0430	.546	.595
Other Personal Prop	5608	4945	.0499	.461	.559
Prop Dam, Prod Liab	2405	1951	.0677	.336	.432
Total	218120	196441	.0358	.737	.612

Note: The number of cases consists of all lawsuits filed in Federal Courts between July 1986 and September 1994. The number of resolved cases consists of the subset of all lawsuits that were resolved (dropped, settled, or tried to a verdict) by September 30, 1994. The trial rate is the fraction of resolved cases that were tried to a verdict by September 30, 1994. The Plaintiff win rate is computed as the ratio of the number of cases where a judgement was entered for the plaintiff divided by the number of cases where a judgement was entered for one party or the other. The plaintiff win rate at trial is computed similarly, but restricting the judgements to those entered after a trial verdict.

TABLE 3
Case Outcomes by Party Identity—Uncensored Cases

Plaintiff	Defendant	Trial Rate (Uncensored)	Plaintiff Win Rate	Plaintiff Win Rate at Trial
Individual	All	.0412 [98833]	.608 [22582]	.586 [3526]
Corporation	All	.0304 [97608]	.836 [29573]	.648 [2486]
All	Individual	.0307 [81762]	.834 [27141]	.645 [2121]
All	Corporate	.0395 [114679]	.632 [25014]	.593 [3891]
Individual	Individual	.0361 [41402]	.724 [10836]	.633 [1263]
Corporation	Individual	.0252 [40360]	.908 [16305]	.662 [858]
Individual	Corporation	.0449 [57431]	.501 [11746]	.560 [2263]
Corporation	Corporation	.0341 [57248]	.748 [13268]	.640 [1628]
All	All	.0358 [196441]	.737 [38449]	.612 [6012]

Note: The numbers in brackets are sample sizes. Information on who won refers to the party for whom a judgement was entered. This is missing for 141,475 of the 196,441 resolved cases. Judgement was entered for both parties in another 2811 cases. These are also treated as missing in these tabulations. Information on who won at trial is missing for 662 of the 7042 cases with trial outcomes. Trial judgements for both parties were entered in another 368 cases.

TABLE 4
Estimates of Probit Models of Probability of Trial

Variable	(1) univariate probit 86-94	(2) univariate probit 86-94	(3) univariate probit 86-90	(4) univariate probit 86-90	(5) bivariate probit 86-94	(6) bivariate probit 86-94
constant	-1.716 (0.02)	-1.604 (0.02)	-1.715 (0.02)	-1.581 (0.03)	-1.754 (0.01)	-1.649 (0.02)
Individual Plaintiff	0.134 (0.01)	0.098 (0.01)	0.135 (0.02)	0.092 (0.02)	0.126 (0.01)	0.089 (0.01)
Individual Defendant	-0.137 (0.02)	-0.025 (0.02)	-0.134 (0.02)	-0.031 (0.02)	-0.151 (0.02)	-0.031 (0.02)
Both Individual	0.026 (0.02)	-0.031 (0.02)	0.016 (0.03)	-0.023 (0.03)	0.046 (0.02)	-0.020 (0.02)
Filing Year Dummies	yes	yes	yes	yes	no	no
Case Category Dummies	no	yes	no	yes	no	yes
ρ					0.458 (0.02)	0.472 (0.02)
log L	-29992.7	-29532.1	-21699.6	-21339.8	-66902.2	-66023.0
n	196441	196441	132086	132086	218120	218120

Note: The numbers in parentheses are asymptotic standard errors. There are 218,120 lawsuits overall and 21,679 censored cases (not resolved by September 30, 1994). The base year, where filing-year dummies are included, is 1986. Controls for cases category, where included, consist of a set of fourteen dummy variables for the fifteen categories listed in table 2. The base category consists of cases where both parties are corporations in insurance cases (where category controls included). The estimates in columns 5 and 6 refer to a bivariate probit model where the second equation (not shown) determines the probability that a case is censored. The censoring function includes the same variables as the trial equation plus a cubic in the time between the date of filing and September 30, 1994. The parameter ρ is the correlation between the errors in the trial and censoring equation.

TABLE 5
Estimates of Log-Normal Competing Risk Model of Case Outcomes

	(1)	(2)	(3)	(4)
	Time to Settlement	Time to Trial	Time to Settlement	Time to Trial
constant	5.481 (0.01)	10.606 (0.05)	5.590 (0.01)	10.281 (0.05)
Individual Plaintiff	0.040 (0.01)	-0.344 (0.03)	0.000 (0.01)	-0.272 (0.03)
Individual Defendant	-0.291 (0.01)	0.316 (0.03)	-0.086 (0.01)	0.038 (0.03)
Both Individual	0.154 (0.01)	0.008 (0.04)	0.020 (0.01)	0.145 (0.05)
Case Category Dummies	no		yes	
σ^2	1.246 (0.00)	2.084 (0.02)	1.226 (0.00)	2.062 (0.02)
ρ	-0.889 (0.01)		-0.889 (0.01)	
log L	-349728.3		-346296.1	
n	218120		218120	

Note: The numbers in parentheses are asymptotic standard errors. There are 218,120 lawsuits overall and 21,679 censored cases (not resolved by September 30, 1994). Durations are measured in days. Controls for cases category, where included, consist of a set of fourteen dummy variables for the fifteen categories listed in table 2. The base category consists of cases where both parties are corporations in insurance cases (where category controls included). The σ^2 parameters are the variances of the errors in the log-duration functions. The parameter ρ is the correlation between the errors in the time-to-settlement and time-to-trial equations.

TABLE 6
Estimates of Simple Probit Models of Probability of Plaintiff Win

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All 86-94	All 86-94	All 86-90	All 86-90	Trial 86-94	Trial 86-94	Trial 86-90	Trial 86-90
Constant	0.820 (0.03)	0.238 (0.03)	0.798 (0.03)	0.210 (0.03)	0.442 (0.06)	0.209 (0.07)	0.407 (0.06)	0.177 (0.08)
Individual Plaintiff	-0.666 (0.02)	-0.559 (0.02)	-0.608 (0.02)	-0.501 (0.02)	-0.207 (0.04)	-0.119 (0.04)	-0.153 (0.05)	-0.065 (0.05)
Individual Defendant	0.661 (0.02)	0.400 (0.02)	0.620 (0.02)	0.372 (0.02)	0.055 (0.05)	0.095 (0.06)	0.104 (0.06)	0.146 (0.06)
Both Individual	-0.079 (0.03)	-0.007 (0.03)	-0.052 (0.03)	-0.018 (0.03)	0.135 (0.07)	0.021 (0.07)	0.066 (0.08)	-0.046 (0.09)
Filing Yr Dummies	yes	yes	yes	yes	yes	yes	yes	yes
Case Cat Dummies	no	yes	no	yes	no	yes	no	yes
log L	-26788.2	-24091.0	-18371.2	-16500.7	-3986.0	-3934.5	-2877.6	-2842.5
n	52155	52155	36611	36611	6012	6012	4368	4368

Note: The numbers in parentheses are asymptotic standard errors. The "All" sample consists of the 52155 resolved cases (filed between 1986 and 1994) where data on whether a judgement was entered for the plaintiff or defendant was recorded. Cases where no judgement was recorded or where judgement was entered for both parties are not included. The subsample in columns 5-8 consists of the 36611 cases that were filed between 1986 and 1990. The "Trial" sample consists of those observations where the judgement was entered after a trial verdict. The base filing year is 1986. Controls for cases category, where included, consist of a set of fourteen dummy variables for the fifteen categories listed in table 2. The base category consists of cases where both parties are corporations in insurance cases (where category controls included).