A Statistical Analysis of Patent Litigation Using Classification Trees

Tammy W. Cowart, Roger Lirely, and Sherry Avery*

I. Introduction

The process of a court making a decision about the outcome of a case has been described as one "marked by reasoning by analogy."[1] Sunstein summarizes the process as follows:

> (1) Some fact pattern A—the "source" case—has certain characteristics, call them X, Y, and Z. (2) Fact pattern B—the "target" case—has characteristics X, Y, and A, or characteristics X, Y, Z, and A. (3) A is treated a certain way in law. (4) Some principle, created or discovered in the process of thinking through A, B and their interrelations, explains why A is treated the way it is. (5) Because of what it shares with A, B should be treated the same way.[2]

Kastellec notes that the process delineated here comprises determining "whether the addition or presence of a certain case fact, *in combination with the absence or presence of clusters of other case facts*, results in a case being deemed in the same class or an altogether different class."[3] The process described by Sunstein and Kastellec is, therefore, a hierarchical process, wherein more weight is given to the presence or addition or absence of some facts or combinations of facts than others.

Scholars who analyze the decision-making process of courts have generally employed one of two broad methodological approaches, qualitative analysis of factors and legal rules used by decision makers (judges or juries) to make decisions or quantitative modeling of the decision-

*Assistant Professor of Business Law, Professor of Accounting, and Assistant Professor of Management, University of Texas at Tyler.

[1] Jonathan P. Kastellec, *The Statistical Analysis of Judicial Decisions and Legal Rules with Classification Trees,* 7 J. Empirical Legal Stud. 202, 205 (2010).

[2] *Id.*, citing Cass R. Sunstein, Legal Reasoning and Political Conflict 65 (New York: Oxford Univ. Press 1996).

[3] Kastellec, *supra* note 1, at 205.

making process, usually employing multivariate regression (logit or probit). The widespread use of both approaches by legal scholars attests to their usefulness in providing valuable insights into the decision-making process of courts.

The beauty of the qualitative approach is that a student of law, in an attempt to map case facts to outcomes, can model the hierarchical nature of legal reasoning, giving greater weight to the presence or addition or absence of one fact or set of facts than to another fact or set of facts. And as Kastellec points out, legal reasoning is not just an additive process but is an interactive one, wherein the presence of fact *q* alone may not matter much in the decision-making process, but the presence of *q* along with the presence of *p* and *r* may matter very much.[4] Kastellec notes, however, legal reasoning suffers from the limitation that it provides no objective quantifiable results.[5] One can only subjectively assess whether the legal reasoning approach will produce analogous outcomes based on analogous facts.

Studies that use logit or probit attempt to quantify the decision-making process, either ascertaining the probability of a certain outcome or assessing the relative influence of various factors on the decision. These studies use quantifiable independent variables, either dichotomous variables (1,0) indicating the presence or absence of factor or continuous variables or integers indicating the magnitude or category of some factor, regressed on a dichotomous dependent variable. The coefficients of the regressors often indicate a correlation between the facts and outcomes and provide insight into the likelihood of an outcome given the presence or magnitude of the associated factor. In addition, goodness of fit statistics provide evidence of how well the specified model fits the data. These methods have enjoyed widespread use by scholars and provide a powerful tool for mapping from case facts to case outcomes.

[4] *Id.* at 206.
[5] *Id.*

Kastellec[6] describes two closely-related assumptions implicit in logit analyses of legal reasoning that may limit their ability to structurally model the relationship between case facts and outcomes. First, in the logit specification, the magnitude and signs of the regression coefficients are additive, meaning the presence or absence of a fact will push or pull a case toward one outcome or the other. The sum of the coefficients indicates the likelihood of a given outcome. Second, because the effect is additive, logit does not analyze interactive effects unless they are specified *ex ante* and based on theory.[7]

II. Classification Trees

Kastellec proposes the use of classification trees to overcome the inherent limitations of both legal reasoning and quantitative analyses.[8] Classification trees offer the opportunity to both model the hierarchical, interactive and dichotomous nature of legal decision-making and empirically test the model's ability to predict case outcomes. This section provides a brief summary of classification trees.

Classification trees are widely used in a number of disciplines including but not limited to medical diagnosis, credit evaluation, and quality management.[9] Classification trees perform two main functions (1) classify cases into groups or (2) predict the value of a target variable based on the values of the predictor values. This paper focuses on the use of classification trees to create decision rules and then predict future events based on the decision patterns. There are often a large number of variables that enter into a decision process. Classification trees are especially helpful in selecting a useful subset of variables from a larger set of variables. In other words, the

---

[6] *Id.*
[7] *Id.*
[8] *Id.*
[9] Chih-Hsuan Wang, Ruey-Shan Guo, Ming-Huang Chiang & Jehn-Yih Wong, *Decision Tree Based Control Chart Pattern Recognition*, INT'L J. PRODUCTION RES. 1, 13 (2007).

method can help identify those variables that have the most impact on the outcome. Classification trees are best used when decision variables can be represented by attribute data and the target decision has discrete output values[10] They provide a clear, documented model of how the decision was made or will be made[11]

Classification trees consist of nodes where decision or chance events occur and arcs (branches) which connect the nodes. The tree provides the framework for the decision process. The classification tree is arranged from bottom to top, with bottom representing the trunk which then branches out based on variables and their interactions to create a branch pattern.[12] Graphically it looks like an upside-down tree. In the context of this research study, the trunk represents the win-loss outcome of a patent case. Each branch is a node which represents a test on an attribute of the variable. The process is terminated when all samples in the leaf belong to the same class.[13] Guh and Shieu[14] recommend that the initial classification tree be created with all variables and values. When the initial tree is built, many branches may include anomalies owing to noise or outliers. This results in over-fitting the data which then requires pruning to remove unnecessary branches.

III. METHODOLOGY

Prior research of patent litigation has examined the various issues involved in patent litigation, questions of patent claim interpretation at the appellate level, trial decisions of district

---

[10] T.M. Mitchell, MACHINE LEARNING 52-80 (McGraw-Hill: New York 1997).
[11] Stuart Eriksen & L. Robin Keller, *Decision Trees, in* ENCYCLOPEDIA OF OPERATIONS RESEARCH AND MANAGEMENT SCIENCE 159-161 (Kluwer Academic Pub., Hingham, MA, 1996).
[12] *Id.*
[13] Wang et al., *supra* note 9.
[14] Ruey-Shiang Guh & Yeou-Ren Shiue, *On-line Identification of Control Chart Patterns Using Self-Organizing Approaches,* 43 INT'L J. OF PRODUCTION RES. 1225 (2005).

courts and the judicial philosophy of Federal Circuit judges.[15] To our knowledge, only one prior study has looked at the effects of the presence or absence of case facts on patent litigation outcomes, that is, who ultimately wins a patent litigation case.[16] Janicke and Ren examined the population of 262 cases for which the Federal Circuit had reached a dispositive decision during the 2002-2004 time period.[17] Using logit, they regressed, on the case outcome, technology area, whether the litigants were individuals or entities, the parties' financial strength (income) level, country of origin and home state of the litigants, whether the trial was by jury or bench, the years of experience of the lead attorneys and whether they were patent attorneys or general practitioners, whether the lead attorney's firm was an intellectual property (IP) boutique firm or a general practice firm, and which judge authored the Federal Circuit opinion. They overcame a convergence problem, anticipated by the requirement that each independent variable have a specified value for both the patentee and the alleged infringer, by dividing the population into two sub-populations, comprising patentee winners and alleged infringer winners. For both sub-populations, only the financial strength of the litigants, whether the trial was by jury or bench and the nature of the lead attorney's firm (IP boutique or general) reached statistical significance. Although there was weak evidence that technology area of the patent may influence the outcome, the sample size was too small, given the number of technology areas identified, to analyze statistically.

The purpose of this study is two-fold. First, we add to the existing literature on patent litigation by extending the work of Janicke and Ren to a later time period and refining their model based on both theoretical and methodological considerations. Second, as suggested by

---

[15] Paul Janicke & LiLan Ren, *Who Wins Patent Infringement Cases?*, 34 AIPLA Q. J. 1 (2006).
[16] *Id.*
[17] *Id.*

Kastellec,[18] we demonstrate the usefulness of classification trees in analyzing the hierarchical nature of legal reasoning in a complex area of litigation and compare them to the results we obtain from logit.

A. Sample Selection and Definition of Variables

We started by identifying all Federal Circuit patent cases for the years 2009-2011 by searching the Lexis/Nexis database[19] using the search term "patent" with at least 5 occurrences, and then limiting the results to Federal Circuit decisions decided after January 1, 2009. The Lexis/Nexis database includes reported and unreported cases, which are updated continually. It is the official publisher in 17 jurisdictions.[20] It was selected for ease of use in searching for dispositive cases. Furthermore, we defined a case as dispositive if the trial court or Federal Circuit court ruled on the validity of a single patent or multiple patents.

1. Winners and Losers

Janicke and Ren[21] defined a patentee winner as a patentee for whom "at least one claim of one patent is finally adjudicated to have been infringed and not invalid or unenforceable…"[22] and an alleged infringer winner as one for whom no claim has these attributes.[23] These definitions are easy to operationalize for cases in which a single patent claim is at issue, or for cases involving multiple patentee claims in which no claim of a patentee is held valid, infringed or enforceable against an alleged infringer. The definitions are somewhat problematic, however,

---

[18] Kastellec, s*upra* note 1.
[19] LexisNexis Legal Research Service, http://law.lexisnexis.com
[20] LexisNexis Legal Research Service, http://law.lexisnexis.com/lexis/Content
[21] Janicke & Ren, *supra* note 15.
[22] *Id* at 4.
[23] *Id.*

when a court holds for the patentee on some claims and for the alleged infringer on others. For example, if a case involves five patent claims, and the court holds that only one claim is valid, infringed and enforceable, the Janicke and Ren[24] definitions would count this as a win for the patentee and a loss for the alleged infringer even though the patentee lost on four of the five claims, and the alleged infringer won on four of the five claims. Additionally, the Janicke and Ren[25] definitions would count the example above as a win for the patentee even if the legal and/or equitable remedies awarded the alleged infringer exceeded those awarded the patentee.

To address these issues, we define winners and losers on the basis of each patent claim asserted in a case, rather than the case as a whole. Theoretically, this approach models the hierarchical decision-making process wherein the adjudicator(s) considers each claim separately and decides, based on the presence or addition or absence of a fact or combination of facts with respect to each claim, whether the claim is valid, infringed and enforceable. For the patentee, each claim that is upheld is a win. For the alleged infringer, each claim that is ruled invalid, not infringed or unenforceable is a win. Defining a win or loss in this manner also facilitates gathering data about the technology associated with a patent,[26] rather than having to make a determination of what technology classification is involved when a case involves multiple patents, representing multiple technology classifications. We gather this data by perusing each case for a determination of the outcome of each claim.

2. Entity Status

[24] *Id.*
[25] *Id.*
[26] See *infra* discussion on technology classification.

Only an individual can invent something, but a patent can be applied for and/or owned by an individual, a company or a group of individuals and/or companies.[27] We use a binary variable to indicate whether the patentee and alleged infringer litigants are individuals or other entities. When there are multiple patentees or alleged infringers and the group includes both individuals and entities, we classify the party as an entity, rather than an individual, under the assumption that the entity, rather than the individual, would use its resources to defend or pursue a claim.

3. Financial Strength

A measure of financial strength should provide a good proxy for the resources a litigant has available to litigate an alleged infringement or defend against same. Janicke and Ren measure the financial strength of litigants by partitioning their data into four levels of net income.[28] Because data are generally unavailable for individuals, they assign individuals to the lowest income category.[29] They use the income of the parent corporation where a subsidiary was the litigant.[30]

Net income, however, is subject to wide fluctuations from year to year. For example, Alcoa, Inc., reported a net loss of $1.1 billion in 2009 but net income of $254 and $611 million in 2010 and 2011, respectively. In each of the three years, their total assets were between $38.5 and $40.1 billion. We select total assets as a proxy for financial strength. We collected data on total assets from *Mergent Online* and the internet. We assigned individuals to the lowest financial strength (total asset) category.

---

[27] Dietmar Harhoff & Markus Reitzig, *Determinants of Opposition against EPO Patent Grants – The Case of Biotechnology and Pharmaceuticals*, Discussion Paper, Institute for Innovation Research and Technology Management (2000), *available at* citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.200.799.
[28] Janicke & Ren, *supra* note 15.
[29] *Id.*
[30] *Id.*

4. Jury or Bench Trial

Intuitively, one would expect that the legal reasoning processes of a person trained and experienced in law (e.g., a judge) might differ from an untrained and/or experienced person or groups of persons (e.g. a jury). In addition, the reasoning process of an individual (e.g., a judge) might differ from the collective reasoning processes of a jury. Prior research supports this intuition. [31] We use a binary variable to identify whether the trial in the court of origin is by bench or jury.

5. Technology Area

Prior research has examined whether patent litigation issues differ across technology types. For example, Allison and Lemley found no differences across nine different technology types in findings of patent invalidity in final district and Federal Circuit court written opinions, but when the technology types were collapsed into three broader classifications, they did find statistically significant differences between the technology types on the grounds for holding a patent invalid.[32] Allison, Lemley and Walker found that software, telecommunications and business methods patents were more frequently litigated than patents of other technologies.[33] Lanjouw and Schankerman also found that the likelihood of a suit being filed and litigated

---

[31] John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q. J. 185 **(**1998)(Juries are likely to favor patentees and are reluctant to question the decision of the PTO); John R. Allison & Mark A. Lemley, *Who's Patenting What? An Empirical Exploration of Patent Prosecution*, 53 VAND. L. REV. 2099 (2000); Janicke & Ren, *supra* note 9(Janicke and Ren found that a jury verdict increased the appellee's chances of winning on appeal); Kimberly A. Moore, *Judges, Juries, and Patent Cases – An Empirical Peek Inside the Black Box,* 99 MICH. L. REV. 365 (2000)(Moore found that patentees are more successful in jury trials than in bench trials, and that juries are more likely to find in favor of the same party when multiple issues are involved).
[32] *See* Allison & Lemley, 26 AIPLA Q. J. 185.
[33] John R. Allison, Mark A. Lemley & Joshua Walker, *Extreme Value or Trolls on Top? The Characteristics of the Most-Litigated Patents,* 158 U. PENN. L. REV. 1 (2009).

differs across technology types.[34] Harhoff and Reitzig found that European patent applications are more likely to be opposed when they involve new technology areas.[35] In a study of new biotechnology companies, Lerner found that companies with high litigation costs avoid patenting in biotechnology subclasses with a history of higher settlement and damage awards than other biotechnology subclasses.[36] Janicke and Ren found weak evidence that litigation win rates differed across technology areas, but convergence problems, resulting from a small sample size and large number of technology areas, precluded their being able to test whether these differences were statistically significant.[37] In summary, the technology of a patent has been a factor in decisions about whether to patent an invention that has higher litigation costs, whether or not to oppose or litigate a patent, and whether or not the patent is valid or infringed.

Researchers have used the United States Patent Classification (USPC)[38], the International Patent Classification (IPC)[39] and/or their own analyses and judgment[40] to identify the technology of a patent. A case can be made for selecting any of these sources. For example, the USPC system first requires the patent applicant to ascertain the technology class and subclass of a patent.[41] Based on the applicant's classification, the patent is assigned to a patent examiner, an expert in that technology field, who, under the supervision of a patent examiner supervisor,

---

[34] Jean O. Lanjouw & Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, 47 J. LAW & ECON. 45 (2004).
[35] Harhoff and Reitzig, *supra* note 27.
[36] Josh Lerner, *Patenting in the Shadow of Competitors*, 38 J. LAW & ECON. 463 (1995).
[37] Janicke & Ren, *supra* note 15.
[38] Lerner, *supra* note 36; *See* Allison & Lemley (1998), *supra* note 31, wherein they used both USPC and experts. (see note 39 *infra*)
[39] Harhoff & Reitzig, *supra* note 27; Lanjouw & Schankerman, *supra* note 34; Jean Lanjouw & Mark Schankerman, *Characteristics of Patent Litigation: a Window on Competition*, 32 RAND J. OF ECON. 129 (2001).
[40] Allison & Lemley (1998), *supra* note 31; Allison, Lemley & Walker, *supra* note 33; Janicke & Ren, *supra* note 15; Sean M. McEldowney, *New Insights on the "Death" of Obviousness: An Empirical Study of District Court Obviousness Opinions*, 2006 STAN. TECH. L. REV. 4 (2006); Arti K. Rai, John R. Allison, & Bhaven N. Sampat, *Frontiers in Empirical Patent Law Scholarship: University Software Ownership and Litigation: A First Examination*, 87 N.C. L. Rev. 1519 (2009).
[41] The United States Patent and Trademark Office, Office of Patent Classification, http://www.uspto.gov/patents/resources/classification/overview.pdf, retrieved March 16, 2012 at 2:18 p.m. CDT.

verifies the classification of the patent.[42] The system is continuously revised, with additional subclasses added as technology advances are made, and patents are reclassified, as needed, when these revisions are made.

As a first step in identifying the technology of patents, we followed the five-class IPC mapping employed by Lanjouw and Schankerman[43] and classified each patent as either (1) drugs and medicine, (2) chemical, (3) electric, (4) mechanical or (5) other. To be more certain that our technology classifications were accurate, we then cross-referenced these classifications for each patent with the USPC.

6. Court of origin

Evidence exists to support the conclusion that patent litigants forum shop.[44] For example, Moore[45] found that both the number of patent cases and the percentage of patent cases relative to other civil cases were larger in a select group of district courts when compared to other districts. She offers several factors a patent litigant might consider when choosing a particular district court as the venue for litigation[46]:

- the knowledge, background, and experience of the judges;
- the judges' previous experience with high technology or patent matters;
- the characteristics, predispositions, and bases of potential jurors;
- the attorney's familiarity with the district and the judges in the district;
- the local rules of the district court;
- the practices of the judges in the district regarding whether they conduct Markman hearings;
- at what point in the litigation the claims will be construed;

---

[42] *Id.*
[43] Lanjouw & Schankerman, *supra* note 34 at 129.
[44] Paul M. Janicke, *Patent Venue and Convenience Transfer: New World or Small Shift?*, 11 N.C. J.L. & TECH. ON. 1 (2009); Kimberly A. Moore, *Forum Shopping in Patent Cases: Does Geographic Choice Affect Innovation?*, 79 N.C. L. REV. 889 (2001); McEldowney, *supra* note 40.
[45] Moore, *supra* note 44.
[46] *Id*.

- the type of evidence the judges will consider in construing the claims;
- the court's docket and its speed in resolving cases;
- the reputation of the parties in the district;
- the convenience for the parties, witnesses and attorneys.[47]

To account for outcomes that might be influenced by regional differences in decision-makers, we identified the six district courts (Big 6) that decided a disproportionately large number of patent cases relative to other district courts. All of the Big 6 districts had over 175 patent cases filed in each year of our study, 2009-2011.[48] Table 1 provides a comparison of the number of patent and other civil cases decided by the Big 6 and the remaining district courts in 2011.[49] We created a discrete variable with a value of 1 through 6 for those cases decided by each of the Big 6 and a value of 0 otherwise.

7. Foreign versus U.S.

Researchers have also examined whether the nationality of the litigants is a factor in litigation decisions. Janicke and Ren[50] found no difference in case win/loss outcomes based on the nationality of the winning party in U.S. cases. Harhoff and Reitzig,[51] however, found differences in the incidence of opposition to patents for U.S. and Japanese companies when contrasted with European companies. We use a binary variable to distinguish U.S. litigants from foreign litigants.

B. Experimental Design

We start by estimating the following logit regression equation:

---

[47] *Id.* at 899-900.
[48] Lex Machina, www.lexmachina.com.
[49] *Id.;* U.S. District Courts-Civil Cases Commenced, by Nature of Suit and District, During the 12-month Period Ending September 30, 2011, *available at* http://www.uscourts.gov/uscourts/Statistics/JudicialBusiness/2011/appendices/C03Sept11.pdf.
[50] Janicke & Ren, s*upra* note 15.
[51] Harhoff & Reitzig, s*upra* note 27.

$$\log\left[\frac{P(x)}{1-P(x)}\right] = \propto + \beta_1\text{COURT} + \beta_2\text{JB} + \beta_3\text{TECH} + \beta_4\text{PENTITY} + \beta_5\text{CNTRY} + \beta_6\text{PASSETS} + \beta_7\text{AIENTITY} + \beta_8\text{ACNTRY} + \beta_9\text{AIASSETS},$$

Where,

COURT	= 1-6 if Big 6 district court, 0 otherwise;
JB	= 1 if jury trial, 0 if bench trial;
TECH	= 1 if technology classification of patent is health or medicine, 2 if chemical, 3 if electrical, 4 if mechanical, 5 otherwise;
PENTITY	= 0 if patentee is an individual, 1 otherwise;
PCNTRY	= 1 if patentee is from U.S., 0 otherwise;
PASSETS	= 1 if patentee's total assets ≥ $1billion, 0 otherwise;
AIENTITY	= 0 if alleged infringer is an individual, 1 otherwise;
AICNTRY	= 1 if alleged infringer is from U.S., 0 otherwise; and

AFIN = 1 if alleged infringers total assets ≥ $1 billion, 0 otherwise.

We are cautious about making statistical inferences from the logit results because the number of observations within certain cells is small.[52] For example, only one patent decision was originally tried in the District Court of New Jersey, and there are only three individual patentees and three individual alleged infringers. We will address this issue in future research by increasing the number of years from which we draw our sample and, thus, will increase our sample size.

Secondly, we enter the nine covariates into the model using the Classification Tree procedure in the SPSS statistical software package to analyze the 88 patent cases. The dependent variable is win/loss for the patent.

We compare the logit and classification tree results by first examining how accurately each procedure predicts the win/loss outcomes of patent decisions. Secondly, we examine the ability of each procedure to model the hierarchical nature of patent decisions.

C. Descriptive Statistics

---

[52] Peduzzi et. al recommend a minimum of ten observations per variable. P. Peduzzi, J. Concato, E. Kemper, T.R. Holford, & A.R. Feinstein, *A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis*, 49 J. OF CLINICAL EPIDEMIOLOGY 1373 (1996).

The final sample comprises 88 patent decisions. Table 2 provides descriptive statistics for the variables. Note that wins outnumber losses by a ratio of nearly 3 to 1. The Big 6 district courts are the original trial court for 42 percent (37 of 88) of the patent decisions. Jury trials outnumber bench trials by a ratio of more than 3 to 1, and electrical and mechanical patents comprise nearly 63 percent of the technology classifications. Nearly all of the litigants, both patentees and alleged infringers, are entities as opposed to individuals and are from the U.S., although there are slightly more alleged infringers from outside the U.S. than patentees from outside the U.S. Interestingly, a large majority of patentees (70 of 88) have total assets of less than $1 billion, whereas a majority of alleged infringers (52 of 88) have total assets in excess of $1 billion.

IV. RESULTS

A. Regression Results

Table 3 contains the logit regression results. The Chi-square statistic of the omnibus test of the model coefficients was 17.22 with nine degrees of freedom, so it is statistically significant at the <.05 level. Only the coefficients of JB (jury or bench trial), TECH (technology classification) and PENTITY (patentee is an entity or an individual) were statistically significant at the <.10 level, but because the model is loglinear and the observations per variable are small, this is of less interest than the signs associated with the covariate coefficients and the odds ratios for each variable. The signs for each coefficient, except AIASSETS, is negative, meaning that a value of >0 for each independent variable decreases the likelihood of a value of 1 (win) for the patent. Specifically this implies that a patent is less likely to win if the case originates in one of the Big 6 district courts, is tried by jury at the district court level, the patentee is an entity, the patentee is from the U.S., the patentee has total assets of >$1 billion, the alleged infringer is an entity, or the alleged infringers is from the U.S. The sole positive signed coefficient (AIASSETS) implies that the patent is more likely to win if the alleged infringer has assets of >$1 billion.

The *Exp(β)* is the odds ratio for each variable and indicates the odds for a win as the covariate increases. An odds ratio of <1 is always associated with a negatively-signed coefficient, indicating that as that covariate increases, the likelihood of a patent win decreases. Of particular interest is that the odds ratio associated with AIASSETS implies that a patent action against an alleged infringer with assets of >$1 billion is 1.3 times more likely to win than a patent action against an alleged infringer with assets <$1 billion.

Table 4 contains the prediction of wins and losses by the logit analysis. Logit correctly predicted 34.8 percent of the wins and 92.3 percent of the losses, for an overall prediction rate of 77.3 percent.

B. Classification Tree Results

The Classification Tree procedure in the SPSS statistical software package was used to analyze the 88 patent cases. The dependent variable is win/loss for the patent. The nine independent variables entered into the model included the court district, jury or bench trial, industry, patentee corporate or individual, patentee country, patentee financial, alleged infringer corporate or individual; alleged infringer country, and alleged infringer financial. The software splits the data into segments that are as homogeneous as possible with respect to the dependent variable to create a classification tree.[53] Of the nine independent variables entered into the model, five variables were identified as having a significant impact on the outcome: court, patentee corporate or individual, industry, patentee financial, jury or bench trial, and alleged

[53] PASW Decision Tree 18, User Guide, page 2, http://www.uky.edu/ComputingCenter/SSTARS/SPSS/18%20Manuals/PASW%20Decision%20Trees%2019.pdf.

infringer country. The software identified seven nodes (Node 0 to 6) and four terminal nodes (Node 2, 4, 5, and 6). The depth of the tree is three, which is the number of levels below the Root node 0. (See Figure 1 for the classification tree.) The tree provides the number and percentage of win/loss for each node.   For example, Node 1 is comprised of the eastern district of Texas, District of New Jersey, and other. In that node 41 (66.1%) of the patent cases are lost and 21 (33.9%) are won.

Figure 2 and Table 5 provide information about the importance of each variable to the patent decision. The selection of district court and technology of the patent are the two biggest predictors of the patent outcome; followed by jury or bench trial, patentee corporate or individual and patentee financial strength.

In addition to identifying the critical variables, the software package also provides statistics about the prediction accuracy of the model (Tables 6 and 7). The risk estimate of .239 indicates that the category predicted by the model is wrong for 23.9% of the cases with a standard error of .045. Conversely, the model predicts the outcome accurately for 76.1% of the cases. The model is most accurate in predicting losses (86.2%) compared to wins (47.8%).

V. Discussion

Both analyses predict patents' outcomes with generally equal success (logit 77.3% correct, decision tree 76.1% correct). Logit appears to be slightly better at predicting losses, 92.3% to 86.2%, but the decision tree analyses appears to be better at predicting wins, 47.8% to 34.8%. A nice feature of the logit analysis is the odds ratio, which allows the researcher to predict the odds of a given outcome for a unit increase in a predictor variable, all other variables held constant.

The advantage of classification tree analysis is that it allows the user to examine the relationship between the variables, not only in determining which variables are most important but which variables in conjunction with other variables are best able to predict an outcome. For example, from Figure 1, we can predict that a chemical or health and medicine patent would have the best chances of winning on appeal if it was originally tried in the Eastern District of Texas, the District Court of New Jersey or another district court outside of the Big 6. A mechanical, electrical or other patent would likely lose on appeal if tried before the bench in those same district courts but would have an increased chance of winning, 33.3% to 6.7%, if originally tried before a jury.

VI. CONCLUSION

The purpose of this study was to add to the body of research of the factors that influence court decisions in patent cases and to illustrate how classification tree analysis can enrich that research. We show that classification tree analysis is at least as accurate as logit analysis in predicting win/loss ratios but provides richer information about the hierarchical nature of patent case decisions and the interaction of variables that influence those decisions. Given the small sample size we use, however, these results should be viewed with caution, especially as they relate to conclusions one might draw about decisions at the district court level or with respect to technology of the patent. Future research with a larger sample may enable us to confirm these preliminary findings and provide information that may be useful to both decision-makers and those who make or influence policy.

Table 1

Number of patent and other civil cases of Big 6 versus other district courts in 2011

District Court	Variable	Patent cases	All private civil cases
E.D. Texas	1	413	2,823
C.D. California	2	316	12,362
N.D. California	3	224	5,835
D. Delaware	4	484	1,209
D. New Jersey	5	178	6,644
N.D. Illinois	6	219	8,330
Other	0	1,757	

Table 2
Descriptive Statistics

Variable	Frequency	Percentage
Outcome		
Win	65	73.9
Lost	23	26.1
Court		
ED Texas	10	58.0
CD California	7	11.4
ND California	2	8.0
D Delaware	13	2.3
D New Jersey	1	14.8
ND Illinois	4	1.1
Other	51	4.5
Jury or Bench Trial		
Jury	67	76.1
Bench	21	23.9
Industry		
Health	18	20.5
Chemical	7	8.0
Electric	34	38.6
Mechanical	21	23.9
Other	8	9.1
Patentee Corporate or Individual		
Individual	3	3.4
Corporate	85	96.6
Patentee Country		
Other	17	19.3
United States	71	80.7
Patentee Financial		
<1 billion	70	79.5
>1 billion	18	20.5
AIF Corporate or Individual		
Individual	3	3.4
Corporate	85	96.6
AIF Country		
Other	24	27.3
United States	64	72.7
AIF Financial		
<1 billion	36	40.9
>1 billion	52	59.1

Table 3

Results of logit regression

Variable	Coefficient	S.E.	Wald statistic	Significance	Exp(β)
COURT	-.324	.206	2.462	.117	.724
JB	-1.789	1.000	3.202	.074	.167
TECH	-.587	.301	3.802	.051	.556
PENTITY	-2.577	1.544	2.783	.095	.076
PCNTRY	-.631	.659	.919	.338	.532
PASSETS	-.616	.746	.681	.409	.540
AIENTITY	-1.705	1.327	1.651	.199	.182
AICNTRY	-.778	.694	1.255	.263	.459
AIASSETS	.251	.582	.186	.666	1.285
Constant	6.369	2.821	5.097	.024	583.556

Table 4
Logit Win/Loss Predicted Classification

	Predicted wins (1)	Predicted losses (0)	Percentage correct
Observed wins (1)	8	15	34.8
Observed losses (0)	5	60	92.3
Overall percentage			77.3

Table 5

Independent Variable Importance

Independent Variable	Importance	Normalized Importance
Industry	.041	100.0%
Court	.039	94.7%
JuryBench	.016	38.2%
PCorpInd	.014	34.0%
ACntry	.008	18.4%
PFinancial	.003	8.1%

Dependent Variable: WinLoss

Table 6

Risk

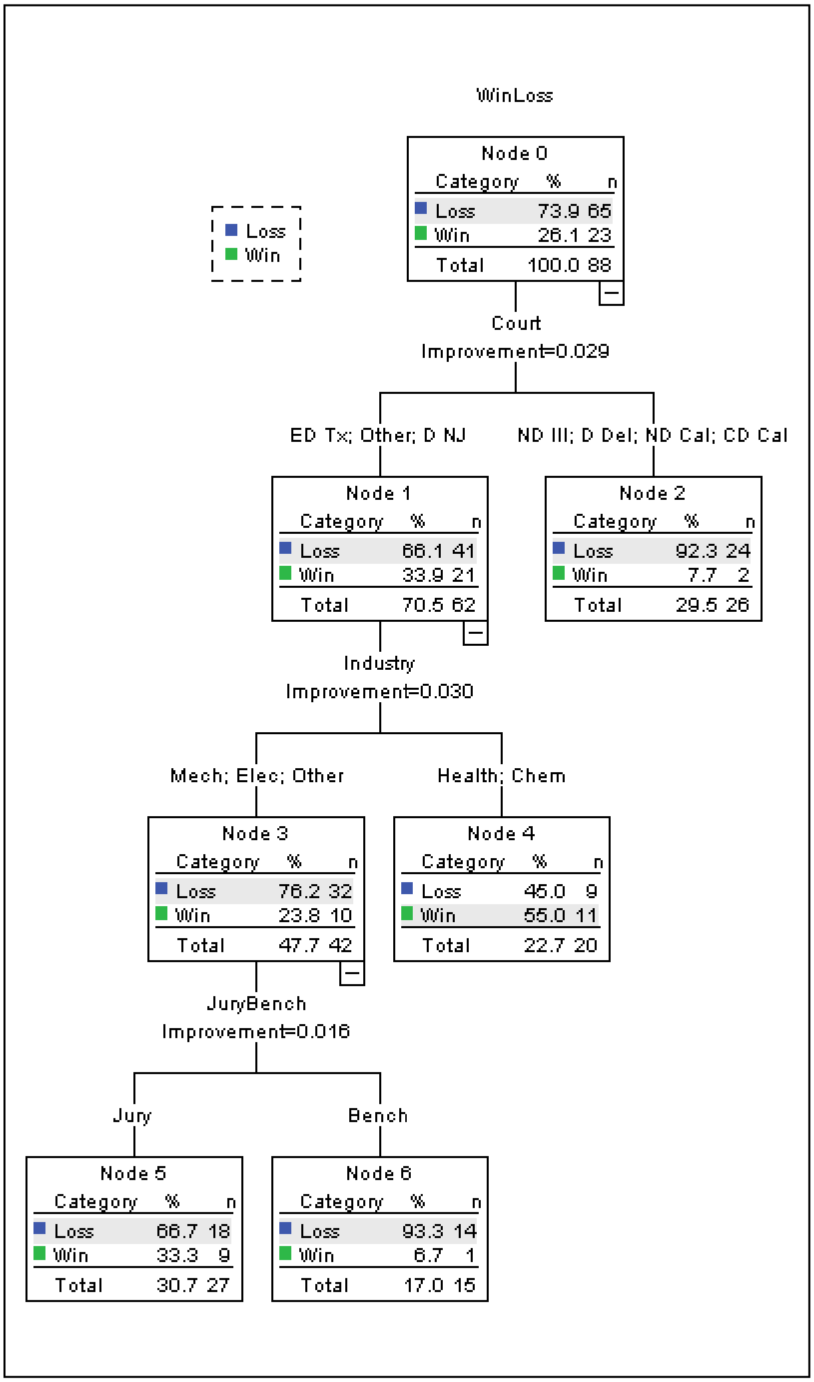
Estimate	Std. Error
.239	.045

Table 7

Classification

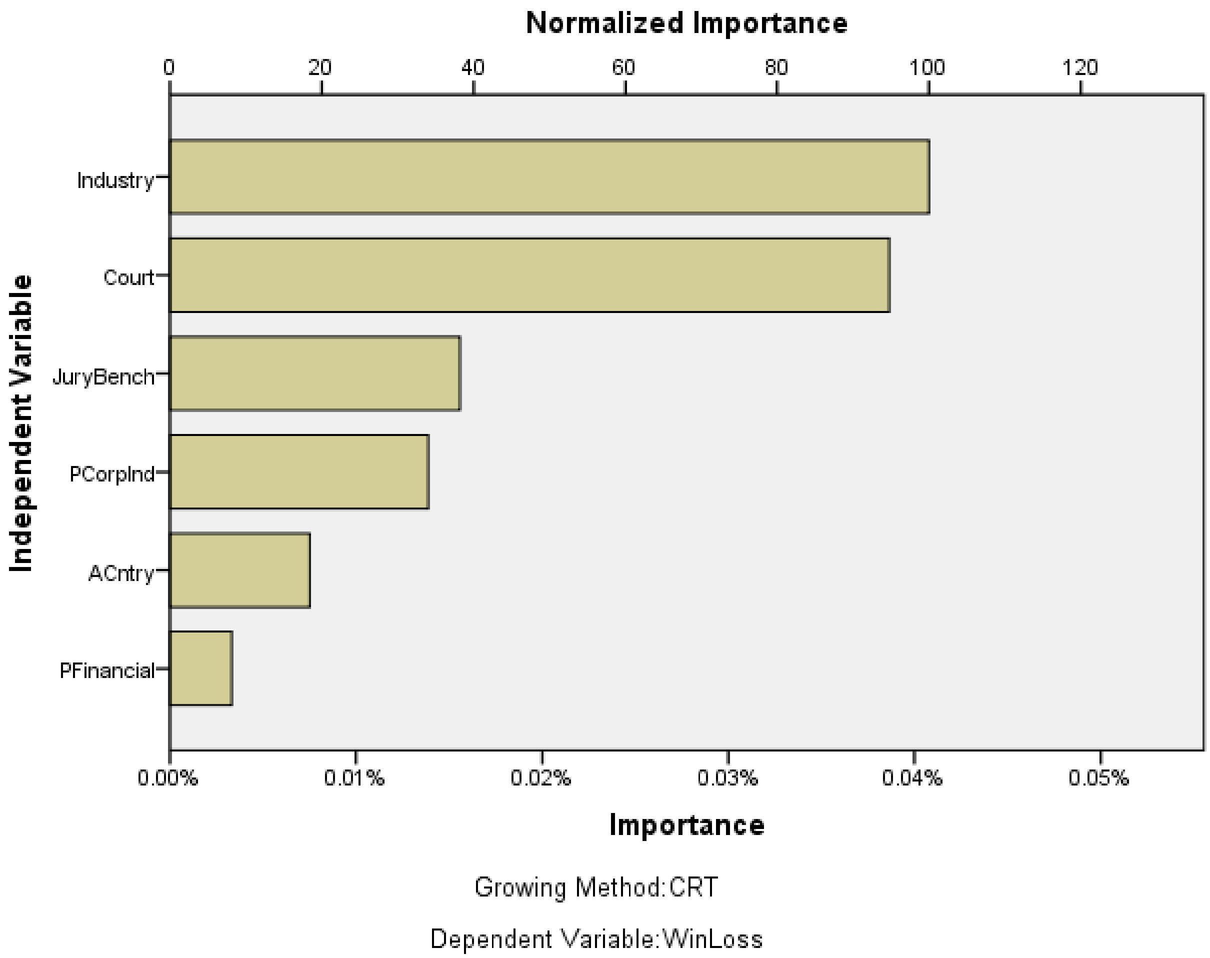
Observed	Predicted		
	Loss	Win	Percent Correct
Loss	56	9	86.2%
Win	12	11	47.8%
Overall Percentage	77.3%	22.7%	76.1%

Figure 1



WinLoss
Loss
Win
Node 0
Category % n
Loss 73.9 65
Win 26.1 23
Total 100.0 88
Court
Improvement=0.029
ED Tx; Other; D NJ
ND Ill; D Del; ND Cal; CD Cal
Node 1
Category % n
Loss 66.1 41
Win 33.9 21
Total 70.5 62
Node 2
Category % n
Loss 92.3 24
Win 7.7 2
Total 29.5 26
Industry
Improvement=0.030
Mech; Elec; Other
Health; Chem
Node 3
Category % n
Loss 76.2 32
Win 23.8 10
Total 47.7 42
Node 4
Category % n
Loss 45.0 9
Win 55.0 11
Total 22.7 20
JuryBench
Improvement=0.016
Jury
Bench
Node 5
Category % n
Loss 66.7 18
Win 33.3 9
Total 30.7 27
Node 6
Category % n
Loss 93.3 14
Win 6.7 1
Total 17.0 15

Figure 2


Normalized Importance
0
20
40
60
80
100
120
Independent Variable
Industry
Court
JuryBench
PCorpInd
ACntry
PFinancial
0.00%
0.01%
0.02%
0.03%
0.04%
0.05%
Importance
Growing Method:CRT
Dependent Variable:WinLoss