Working Paper 96-21
Economics Series 11
April 1996

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-1) 624-9875

THE DISTRIBUTION OF SENTENCES IN TAX-RELATED CASES: EVIDENCE FROM SPANISH COURTS OF APPEALS

Javier Estrada and Santos Pastor *

Abstract

The distribution of sentences in tax-related cases in Spain shows that the government tends to lose more often in this type of cases than in any other type of administrative cases; it also shows that such distribution varies widely across the type of taxes and other variables. Our purpose is thus twofold: First, we attempt to identify the factors that explain the result of tax-related cases; then, we use those factors to build a model to forecast the government's probability of success in this type of cases.

Keywords
Litigation. Taxation. Statistics and the law.

JEL Number: K41

* Departamento de Economía de la Empresa y Departamento de Economía, Universidad Carlos III de Madrid. EMAILS: <estrada@eco.uc3m.es> and <pastor@eco.uc3m.es>. We would like to thank Gabriel Elorriaga and Rich Sicotte for helpful comments. We gratefully acknowledge financial support from the Instituto de Estudios Fiscales. The views expressed below and any errors that may remain are entirely our own.

I- INTRODUCTION

The distribution of sentences on tax-related cases in Spain is significantly different from that of other administrative cases. As argued in Pastor (1993), one of its distinctive aspects is that the government tends to lose more often in tax-related cases than in any other type of cases. Another is that the distribution of sentences varies widely depending on the type of taxes. Thus, the purpose of this article is twofold: First, we attempt to identify the factors that determine the result of tax-related cases; and, second, we use those factors to build a model with the purpose of forecasting the government's probability of success in this type of cases.

In a seminal article, Priest and Klein (1984) argued that the cases that go to trial are those in which there is relatively more uncertainty about their outcome. This divergent-expectations theory, which follows from the fact that cases that clearly favor either the plaintiff or the defendant are settled out of court, implies that a plaintiff (or a defendant) is expected to win at trial with a probability of 50%. This theory further argues that deviations from the 50% predicted rate can be explained by the stakes of the case (for example, in cases of medical malpractice, product liability, or workplace injuries), among other factors.

The divergent-expectations theory has not been devoid of critics; both theoretical arguments and empirical evidence have been put forward against it. From a theoretical point of view, Hylton (1993) argues that a party's proportion of victories could be higher than 50% if this party has an informational advantage or is favored by the burden of proof, among other reasons. From an empirical point of view, on the other hand, Eisenberg (1990) reports that the proportion of cases won by plaintiffs ranges between 52% and 84% in contract cases and between 25% and 60% in personal-injury cases.[1]

Our study uses a sample of over 1,000 tax-related cases decided by the Spanish Courts of Appeals during the first quarter of 1992. With these data we attempt to show that, due to the characteristics of the area of the law we consider, there is no reason to expect a specific rate of success for the government or the plaintiffs like, for example, the 50% rate predicted Priest and Klein (1984). Although in principle any distribution of court outcomes

[1] Long before Eisenberg (1990) and Hylton (1993), and even before Priest and Klein (1984), Galanter (1974) suggested that a plaintiff's probability of prevailing in court can be explained by the differential stakes of the parties, as well as by the nature of the parties (that is, by considering whether the plaintiff and the defendant are one-shot players or repeat players). Relevant to this discussion, see also Shavell (1995)

is possible in our sample, we will argue that there exist factors (for example, budget constraints, risk aversion, and variables affecting threshold probabilities) that have the impact of decreasing the dispersion of this distribution. In addition, we generate results that may lead to the identification, and hopefully correction, of problems in the system. To the best of our knowledge, this is the first time that our data is analyzed with the approach we follow in this article.

The rest of the article is organized as follows. In part II, we consider a plaintiff's decision about whether to try a tax-related case under two possible scenarios, namely, one in which the type of tax allows the option of settling and one in which it does not. In part III, we derive and estimate a model with two purposes: First, to identify the factors that affect the government's probability of success in tax-related cases; and, second, to forecast the government's probability of success in this type of cases. Finally, in part IV, we summarize the main implications of our analysis. An appendix containing several tables concludes the article.

II- THE PLAINTIFF'S DECISION

We consider in this part two different frameworks, both stemming from the characteristics of our data. Although the data is discussed in more detail below, it suffices to note at this point that the cases in our sample can be classified into those in which settlement is possible and those in which it is not. In those cases in which settlement is possible, the plaintiff may choose to pay a fine and settle the case in exchange for a 50% reduction in the due tax. In those cases in which settlement is not possible, on the other hand, a plaintiff's claim goes directly to court.[2] We analyze in the next two sections the strategic behavior of plaintiffs in cases in which settlement is not possible, and then in cases in which it is.

1.- The Plaintiff's Decision When Settlement Is Not Possible

In those cases in which settlement is not possible, the critical decision taken by a plaintiff is whether or not to sue. As an example, consider the capital gains tax, one of the frequent sources of litigation. The process starts when the government sends a taxpayer a

[2] To be sure, this classification of cases into those in which settlement is possible and those in which it is not is not always clear; but it is in fact the case that settlements are more frequent in some type of cases than in others. To illustrate, settlements tend to be common in income tax cases taxes and rare in capital gains cases.

tax form stating an amount due; the taxpayer can either pay the stated amount or refuse to pay and file a claim. In the latter case, since settlement is not possible, then the claim always leads to litigation; hence, the plaintiff must decide the amount of resources to be allocated to trying the case. Finally, if the plaintiff loses in court, he must decide whether to appeal the decision, and, if he does appeal, must decide the amount of resources to be allocated to this second stage of the litigation process.

A simple model. In order to formalize this discussion, let a plaintiff receive a tax bill sent by the government for the amount *S*, which the plaintiff thinks is incorrect. Let p_p be the plaintiff's probability of winning a case, and *C* be his litigation costs under the American rule. Thus, a rational (risk-neutral) plaintiff will sue if and only if $p_p * S > C$; that is, if the plaintiff's expected gain from suing (not paying the amount *S*) is larger than his litigation costs. Or, put differently, he will sue as long as long as the probability of winning the case is larger than a given threshold; that is, as long as

$$p_p > C / S . \qquad (1)$$

Thus, if the plaintiff's litigation costs were, say, 10% of the stakes, the plaintiff will sue even if the probability of winning is as low as, say, 15%.

Economies of scale in litigation. It is generally the case that, due to the existence of economies of scale in litigation, the ratio of litigation costs to stakes (*C/S*) decreases as the stakes increase. For example, if trying a case in which the stakes are $1 million costs $10,000, it is generally not the case that trying a case in which the stakes are $6 million costs $60,000; in general, it will cost less than that. As a result, larger stakes imply a reduction in the threshold *C/S* and a subsequent increase in the plaintiff's incentive to sue. Note, however, that as long as larger stakes do not imply larger merits (and there is no reason why they should), neither they imply a change in the plaintiff's probability of success. Therefore, the larger the stakes, the larger the government's probability of success at trial.

To illustrate, suppose a plaintiff is considering two suits: In the first, the stakes are $100,000, the plaintiff's litigation costs are $20,000, and his probability of winning is 60%; in the second, the stakes are $500,000, the plaintiff's litigation costs are $100,000, and his probability of winning is 10%. Note that in both cases the plaintiff's litigation costs are 20% of the stakes. Note, further, that in the absence of economies of scale (or if the cases are considered in isolation), a rational plaintiff should not try the second case; this is due to the fact that, according to (1), a case should be tried when $p_p > C/S$, which is not the case in the example (p_p=10% < C/S=20%). However, if economies of scale bring the litigation costs of the second case down to $25,000, then a rational plaintiff will sue (because p_p=10% > C/S=5%). In short, the existence of economies of scale in litigation implies that a larger proportion of "bad" cases (that is, cases in which the plaintiff's probability of success is low) will go to court, thus increasing the government's probability of success.

Litigation costs. As discussed above, besides the decision of whether or not to sue, a plaintiff that decided to sue must decide the amount of resources to be spent in litigation. This decision is important because the probability of winning a case obviously depends on the resources allocated to the case by the plaintiff and the government. To formalize, let R_p and R_g be the amount of resources allocated to the case by the plaintiff and the government, respectively; further, let p_g be the probability that the government prevails at trail, and F be all the other factors that affect the probability of success of the plaintiff and the government. It is then the case that $p_p=p_p(R_p,R_g,F)$ and $p_g=p_g(R_p,R_g,F)$, such that $\partial p_p/\partial R_p>0$, $\partial p_p/\partial R_g<0$, $\partial p_g/\partial R_p<0$, and $\partial p_g/\partial R_g>0$.

If the government ranked cases by their importance (measured by stakes, impact on reputation, and the like), and then attempted to maximize the number of wins subject to a budget constraint, the government's optimal choice would maximize social welfare; that is, the sentences imposed on the plaintiffs would equal the plaintiffs' expected gains from cheating on the government. However, the government's usual decision of allocating an approximately equal amount of resources to each case does not maximize social welfare; instead, it wastes resources and achieves a lower-than-optimal level of deterrence.[3] As a result, the government's failure to optimize the amount of resources to be allocated to each case tends to decrease the government's probability of success at trial.

The role of intermediate courts. One last issue we address in this section is the role performed by the Regional Tax Courts and the Central Tax Court. In particular, we argue that the larger the proportion of cases that go through these intermediate courts, the larger the government's probability of success should be. This follows from the fact that the Central Tax Court and the Regional Tax Courts perform the function of deciding the "best"

---

[3] Further, given this behavior, R_p should be increasing in S; see Pastor (1994).

cases (that is, the cases with a large probability of a plaintiff's win) in favor of the plaintiffs; as a result, the cases that reach the Courts of Appeal (which are the ones in our sample) have a large proportion of "bad" cases.

To illustrate, assume that there are 50 "good" cases and 50 "bad" cases. Assume, further, that a Regional Tax Court decides 25 of the "good" cases in favor of the plaintiffs, the other 25 "good" cases in favor of the government, and the 50 "bad" cases also in favor of the government. Note that, in the original sample of 100 cases, the government was expected to win 50% of the cases; however, out of the 75 cases that reach the Court of Appeal (after having gone through the filter of the Regional Tax Court), the government is expected to win 50 cases; that is, 75%. Thus, as argued above, the impact of having cases go through the Central Tax Court or the Regional Tax Courts is that of increasing the government's probability of success at trial.

There is, however, one opposing effect in those cases that go through the Central Administrative Court. Typically, the stakes involved in these cases are large, thus inducing plaintiffs to spend relatively more resources in litigation; hence, the government's probability of success at trial should decrease.

2.- The Plaintiff's Decision When Settlement Is Possible

In those the cases in which settlement is possible, the plaintiff may agree to pay a fine and settle the case, thus being entitled to a 50% reduction in the tax due. To formalize this alternative situation, let P be the penalty and R the accrued interest on the unpaid amount, and define $Q=S+P+R$. The payoffs of a potential plaintiff when he chooses to settle and to sue are respectively given by $Q-(1/2)S$ and $(1-p_p)Q+C$; hence, a rational (risk-neutral) plaintiff will sue if and only if $(1-p_p)Q+C < Q-(1/2)S$. Or, put differently, he will sue as long as the probability of winning the case is larger than a given threshold; that is, as long as

$$p_p > (1/2)(S/Q) + C/Q. \qquad (2)$$

To illustrate, if the amount at stake is \$400,000, the penalty is \$400,000, the accrued interest is \$200,000, and the litigation costs are \$100,000, a rational plaintiff will sue only if his probability of winning is larger than 30%.

Recall that, in the absence of penalties (and interest payments), a rational plaintiff sues as long as $p_p>C/Q$; hence, a plaintiff facing the numbers of our last example would sue as long as he perceives the probability of winning larger than 10%. It is thus interesting to note that, if a plaintiff thinks that his chances of winning the case are 20%, he would sue if the issue at stake (that is, the type of tax) does not allow the option of settling the case, but would not sue if the issue at stake makes a settlement possible. In other words, the existence of the settlement option increases the probability that a plaintiff does *not* go to court with a "bad" case, thus decreasing the government's probability of success.

In sum, a comparison of the distribution of cases in which settlement is possible and the distribution of cases in which settlement is not possible should exhibit two main differences: First, the former should have fewer cases; and, second, the former should have a smaller proportion of government's victories.

3.- Other Factors Affecting the Plaintiff's Decision

Besides the factors analyzed above (basically, stakes, litigation costs, penalties, and probabilities of success), there exist several other factors that determine the decision of plaintiffs, and, therefore, the characteristics of the distribution of sentences. Two of those factors are the merits of a plaintiff's case and the plaintiff's informational advantage; these typically refer to the plaintiff's ability to prove his case, as well as to the fact that he knows (and generally the government ignores) whether or not he is guilty.

Risk aversion also plays an important role in a plaintiff's decision of whether or not to sue; see Perloff, Rubinfeld, and Ruud (1993). In particular, the higher the plaintiff's risk aversion, the higher the threshold; that is, the higher the minimum probability of winning that induces a plaintiff to go to court. If it is the case that, as is plausible to assume, plaintiffs are generally more risk averse than the government, then the impact of risk aversion would be that of decreasing the government's probability of success; this follows from the fact that, the higher the level of risk aversion, the "better" a case must be for a plaintiff to take it to court.

Plaintiffs (as well as the government) also take into account issues of reputation when deciding whether to try a case. Plaintiffs, in particular, may be concerned that the very fact of suing may increase the probability of being audited in the future. The government, on the other hand, may decide to try some apparently-unimportant cases (in terms, for example, of their stakes) only to send the signal that even minor transgressions of the law will not be tolerated.

Finally, and opposing the effect discussed immediately above, budget constraints impose limitations on the number of cases the government can litigate and on the amount of resources it can allocate to each case. Thus, budget constraints may prevent the government from trying a case in which the probability of winning is very low, thus inducing some plaintiffs to sue in circumstances in which, budget constraints notwithstanding, they would not.

4.- A Synthesis

We briefly summarize in this section the main issues discussed in this part. We started by classifying cases into those in which settlement is possible and those in which it is not. We then argued that the existence of the settlement option has a negative impact on the government's probability of success.

We emphasized that factors that affect thresholds (namely, stakes, litigation costs, and penalties) are important factors determining the plaintiff's decision about whether or not to sue. More precisely, we argued that any factor that lowers the threshold but leaves the plaintiff's probability of winning unchanged increases the plaintiff's incentives to sue, thus increasing the proportion of "bad" cases in the sample, and, as a result, the government's probability of success.

Other factors affecting the government's probability of success include the merits of the case, the plaintiff's informational advantage, the plaintiffs' risk aversion, the government's budget constraint, and issues of reputation. In particular, risk averse plaintiffs will tend to try relatively "good" cases, thus decreasing the government's probability of success.

III- EMPIRICAL RESULTS

We present in this part the results of our empirical analysis. We start by introducing the econometric framework, then we move to discuss the data and the variables to be considered in the model, and then we state our basic hypotheses. Subsequently, we present the results of our initial estimations and of some refinements, we perform out-of-sample forecasting in order to evaluate the predictive ability of the model, and, finally, we illustrate a possible use of the models with two examples.

1.- The Econometric Framework

Let an index summarizing the government's ability to prove the *i*th case (M_{gi}) and another summarizing the plaintiff's ability to prove the same case (M_{pi}) be expressed as a linear function of k variables; that is,

$$M_{gi} = \sum_{j=1}^{k+1} a_j x_{ij} + v_i \tag{3}$$

$$M_{pi} = \sum_{j=1}^{k+1} b_j x_{ij} + w_i \,, \tag{4}$$

where $(a_1,...,a_{k+1})$ and $(b_1,...,b_{k+1})$ are vectors of coefficients, $(x_1,...,x_{k+1})$ is a vector of explanatory variables such that x_1=1, v and w are normally-distributed (0-mean) random variables, and n is the number of observations (cases) in the sample. Thus, subtracting (4) from (3) yields the government's relative advantage to prove the ith case (M_i^*), which is given by

$$M_i^* = M_{gi} - M_{pi} = \sum_{j=1}^{k+1}(a_j - b_j)x_{ij} + (v_i - w_i) = \sum_{j=1}^{k+1} \beta_j x_{ij} + u_{ij} = I_i + u_i \,, \tag{5}$$

where $\beta_j=(a_j-b_j)$, $u_i=(v_i-w_i)$, and $I_i=\sum_{j=1}^{k+1} \beta_j x_{ij}$.

Let a court's decision of the ith case (D_i) be represented by a dummy variable such that D_i=1 represents a situation in which the court decides in favor of the government, and D_i=0 one in which the court decides in favor of the plaintiff. The government will be successful (D_i=1) only if it presents a better case than the plaintiff; that is, only if $M_{gi}>M_{pi}$. It thus follows from (5) that the government will win the ith case only when $M_i^*>0$, which, in turn, implies that $I_i+u_i>0$. Therefore, the probability that the government wins the ith case, $p_g=p_g(D_i=1)=p_g(M_i^*>0)$, is given by

$$p_g = p_g(I_i + u_i > 0) = 1 - p_g(u_i < -I_i) = 1 - F(-I_i) = F(I_i) = \int_{-\infty}^{I_i} f(u_i)du_i \,, \tag{6}$$

where F and f indicate the cumulative distribution function and the probability distribution of u, respectively. In words, equation (6) states that the probability that the government wins the ith case is equal to the probability that that the government presents a better case than the plaintiff's, which, in turn, is given by the cumulative distribution function of the random variable u.

Finally, since v and w are normally distributed so is u, thus enabling the analytical model to collapse into a probit model; that is, a model in which the government's probability of winning the ith case is given by

$$p_g(D_i = 1) = (2\pi)^{-1/2} \int_{-\infty}^{I_i} e^{-\frac{u_i^2}{2}} du_i = \Phi(I_i), \tag{7}$$

where Φ denotes the cumulative standard Normal distribution function. Equation (7) thus summarizes the econometric framework and, after introducing the explanatory variables and estimating the model, is the equation that will be used to forecast the government's probability of success in the tax-related cases it litigates.

2.- The Data and the Basic Model

The purpose of our estimations is, as discussed above, twofold: First, we attempt to identify the factors that determine the government's probability of success; second, we attempt to use those factors to predict such probability in the tax-related cases it litigates.

The sample under consideration consists of a set of 1,208 cases decided during the first quarter of 1992 by the Courts of Appeal in the cities of Madrid, Barcelona, Burgos, Valladolid, La Coruña, Valencia, Granada, Málaga, Seville, Las Palmas (Gran Canaria), and Santa Cruz (Tenerife). The dependent variable under consideration (SNTNC) represents a Court of Appeal's sentence on a tax-related case in which the government was a defendant. By definition, this variable takes a value of 1 when the government wins the case, and a value of 0 when the government loses the case.

Our basic hypothesis is that the government's probability of success is determined by eight types of variables, namely, the *type of tax* involved in the case, the *stakes* of the case, the identity of the *plaintiff* and that of the *defendant*, the existence of a previous sentence by *intermediate courts*, the *length* of the process, the *amount of resources* spent by the plaintiff, and the (appeals) *court* in which the case is tried. We now move to describe each of these variables in more detail.

We consider five different *types of taxes* in the model, namely, capital gains taxes (TCG), service fees (TSF), income taxes (TI), capital transactions taxes (TCT), and real estate taxes (TRE). We classify the *stakes*, on the other hand, in three categories, namely, those smaller than 350,000 pesetas (SS350), those between 350,000 pesetas and 1 million pesetas (S3501M), and those larger than 1 million pesetas (SL1M).[4] All variables representing taxes and stakes are dummy variables.

[4] The base category in the stakes variable is given by those awards not determined by the courts.

Among all possible *plaintiffs*, we focus our attention on two groups, namely, private firms (PPF) and individuals (PI). In terms of *defendants*, on the other hand, we initially bundle in one variable (DLAC) the local governments, the autonomous governments, and the councils. Subsequent to the initial phase of estimation, we disaggregate the variable DLAC into three variables: one for the local governments (DLG), one for the autonomous governments (DAG), and one for the councils (DC). In addition, to assess the impact of *intermediate courts*, we include one variable representing cases that go through a Regional Tax Court (ICRTC) and another representing cases that go through the Central Tax Court (ICCTC). All the variables representing plaintiffs, defendants, and intermediate courts are dummy variables.

In order to account for the *length* of the process, we incorporate a variable that measures the number of days between the filing of the case and the sentence (LFS), which is the only non-dummy variable of the model. Our proxy for the *amount of resources* spent by plaintiffs, on the other hand, is given by the legal team that represents those plaintiffs; we thus include three dummy variables: one for those plaintiffs represented by a legal administrator (RLA), one for those represented by an attorney (RA), and one for those represented by both a legal administrator and an attorney (RLAA).[5]

Finally, we incorporate into the model ten variables in order to assess the relationship between the result of the cases tried by the government and the *court* in which those cases are tried. We thus include dummy variables representing courts in the cities of Madrid (CMDRD), Barcelona (CBRCLN), Burgos (CBRGS), Valladolid (CVLLDLD), La Coruña (CLCRN), Valencia (CVLNC), Granada (CGRND), Málaga (CMLG), Seville (CSVLL), Las Palmas de Gran Canaria (CLPGC), and Santa Cruz de Tenerife (CSCT). All these variables are incorporated into the model in a stage subsequent to the initial phase of estimation.

Some summary statistics of all the variables defined in this section are reported in Table A1, in part 1 of the appendix. We now move to discuss the expected impact of these variables on the government's probability of success.

[5] The base category of the representation variable is given by those plaintiffs that represent themselves in court.

3.- Expected Signs of the Coefficients

We present in this section, based on the theoretical arguments presented in the previous part, our expectations about the signs of the coefficients of the variables introduced in the last section. There exist, however, two shortcomings: First, the direct impact of one variable on the government's probability of success may also generate indirect effects of opposite sign; and, second, equally plausible alternative explanations may imply, in several cases, opposing expected signs for a given variable. In both circumstances, we rely on the empirical analysis to determine which effect predominates.

The difference among the *types of taxes* is relevant for at least three reasons. First, because the type of tax determines whether settling the case is a feasible option; as seen above, the existence of the settlement option increases the probability that a plaintiff does not go to court with a "bad" case, thus decreasing the government's probability of success. Second, because the type of tax determines whether the case goes through an intermediate court or goes directly to the Courts of Appeal; as was also seen above, an intermediate court that filters "good" cases has a positive impact on the government's probability of success. Third, because the type of tax is correlated with the identity of the plaintiff and the defendant. As a result of these three effects, the expected sign of most tax variables is ambiguous.

To illustrate this ambiguity, consider the capital gains tax. On the one hand, in conflicts that stem from this tax, settlement is not possible; this, as argued above, increases the government's probability of success. On the other hand, conflicts that stem from this tax do not go through the filter of the Central Tax Court or a Regional Tax Court; this, as also argued above, decreases the government's probability of success. Hence, these two opposing arguments force us to rely on the data to determine the empirical impact of trying a capital gains case on the government's probability of success. The same two conflicting arguments (and implication) hold for the service fees tax and for the real estate tax.

Unlike conflicts that stem from the capital gains tax, the service fees tax, or the real estate tax, in conflicts that stem from the income tax, settlement is usually possible; in addition, these conflicts usually go through the filter of the Central Tax Court or a Regional Tax Court. Thus, though the existence of the settlement option decreases the government's probability of success, the filter of an intermediate court increases such probability. Hence, the expected sign for the income tax variable is also ambiguous.

There is, however, no ambiguity about the expected sign of the capital transactions tax. Conflicts that stem from this tax do not usually allow the settlement possibility, thus increasing the government's probability of success. On the other hand, conflicts that stem from this tax usually go through the filter of the Central Tax Court or a Regional Tax Court, thus increasing again the government's probability of success. Hence, we would expect the variable representing transaction taxes to have a positive sign.

The *stakes* of the case have, in theory, an ambiguous impact on the government's probability of success. On the one hand, higher stakes should induce both the plaintiff and the government to spend more in the case. However, as discussed above, the government does not perform (at least as well as plaintiffs do) an optimization of the resources to be allocated to each case; rather, it tends to distribute resources equally among cases. Thus, higher stakes would tend to decrease the government's probability of success. On the other hand, it follows from equations (1) and (2), and the argument related to the existence of economies of scale in litigation, that an increase in the stakes lowers the threshold, thus increasing the probability that a plaintiff does not go to court with a "bad" case, thus increasing the government's probability of success. Hence, the impact of the stakes on such probability is also ambiguous.

In terms of *plaintiffs*, we would expect individuals, and, in particular, private firms, to have a negative impact in the government's probability of success. This is due to the fact that, in relative terms, these parties tend to allocate more resources than others to litigate their cases. In terms of *intermediate courts*, on the other hand, we already argued that the larger the number of cases that go through the filter of either the Central Tax Court or a Regional Tax Court, the larger the probability of a government's win should be. Recall, however, that the relatively larger stakes of the cases that go through the Central Tax Court induce plaintiffs to spend more, thus decreasing the government's probability of success. In other words, the two opposing effects imply ambiguity about the impact of the Central Tax Court on such probability.

Our proxy for *the amount of resources* spent by plaintiffs at litigation is the legal team that they have chosen to represent them. In this regard, we would expect all three variables to have a negative impact on the government's probability of success. Finally, we have no hypotheses concerning the impact of trying a case in any given city; we will thus rely on our data to tell us whether there is any relationship between the government's probability of success and the cities in which the trials take place.

4.- Initial Estimation

We start our empirical analysis by estimating a probit model with seventeen explanatory variables (*Model 1*); in this first model, we neither disaggregate the variable DLAC nor we include the court variables. We thus estimate a model with five variables representing taxes (TCG, TSF, TI, TCT, and TRE), three variables representing stakes (SS350, S3501M, and SL1M), two variables representing plaintiffs (PPF and PI), one variable representing defendants (DLAC), two variables representing intermediate courts (ICRTC and ICCTC), one variable representing the length of the trial (LFS), and three variables representing the amount of resources spent by plaintiffs (RLA, RA, and RLAA). The results obtained for this first estimation are reported in Table A2, in part 2 of the appendix.

Table A2 shows that, perhaps as expected in a model with seventeen explanatory variables, and as also expected from our theoretical arguments, not all the variables are significant. In particular, none of the variables representing stakes, the amount of resources spent by plaintiffs, the type of defendant (DLAC), and the length of the trial are significant. Nor are significant two variables representing taxes (TSF and TCT) and one variable representing intermediate courts (ICRTC). Thus, our first estimation shows that the government's probability of success is significantly affected by three variables representing taxes (TCG, TI, and TRE), by both variables representing plaintiffs (PPF and PI), and by one variable representing intermediate courts (ICCTC). The fit of the model is not impressive in terms of the Cragg-Uhler R^2 (.1069),[6] although it is not clear that the R^2 in probit models can be interpreted as the proportion of variability in the dependent variable explained by the independent variables.[7] Finally, Model 1 has a reasonably-good in-sample prediction success, achieving 61.17% of correct predictions.

[6] The upper bound of the R^2 in probit models is not necessarily 1; such upper bound is actually a function of the value of the log-likelihood function for the constant (LLFC) and the number of observations in the sample (n). More precisely, this relationship is given by $R^2 \leq 1\text{-}\exp(LLFC)^{2/n}$; see Maddala (1983). Such expression enables us to determine that, in Model 1, the upper bound for the R^2 is in fact 1.

[7] In addition, as shown by Morrison (1972), it is usual to obtain low R^2s in models with qualitative dependent variables, although such low values do not necessarily imply that the estimated model is not good. See, however, Goldberger (1973).

The large number of variables in Model 1 (and the theoretical relationships among them) makes the existence of multicollinearity very likely. As a result, we decided to eliminate nonsignificant variables sequentially, as opposed to eliminate them all at once. We thus started the process by dropping the least-significant variable, re-estimating the model, eliminating the new least-significant variable, re-estimating again the model, and so forth, finally converging to a model with all significant variables. Such an iterative procedure led us to a model with seven explanatory variables (*Model 2*), namely, three variables representing taxes (TCG, TI, and TRE), two variables representing plaintiffs (PPF and PI), one variable representing defendants (DLAC), and one variable representing intermediate courts (ICCTC). The results obtained for Model 2 are reported in Table A3, in part 2 of the appendix.

Table A3 shows that the seven explanatory variables of Model 2 are significant. More precisely, it shows that (everything else equal) the government is significantly more likely to win cases about capital gains taxes and real estate taxes, and significantly more likely to lose cases about income taxes, cases in which the plaintiffs are private firms or individuals, cases in which the defendant is the local government, the autonomous government, or the councils, and cases that go through the Central Tax Court. The Cragg-Uhler R^2 of Model 2 is slightly lower than that of Model 1, but a likelihood-ratio test (LRT) enables us to determine that the ten nonsignificant variables can be dropped from the model without a significant decrease in the likelihood function.[8] Finally, the in-sample prediction success of Model 2 is slightly better than that of Model 1, achieving a 61.42% of correct predictions.[9]

5.- Refinements

We extend in this section Model 2 in two ways: First, we disaggregate the variable DLAC, and, second, we introduce the court variables in order to determine whether there exists any relationship between the government's probability of success and the courts in which the cases are litigated.

[8] $LRT = 2*(789.08-783.88) = 10.4 < \chi^2_{10,.05} = 18.31$.

[9] Note that Model 2 seems to be "pessimistic" in its predictions. This follows from the fact that out of 466 incorrect predictions, in 366 (78.54%) the model predicts that the government will lose when the government actually won the case.

We begin by disaggregating the variable DLAC into its three components (DLG, DAG, and DC) with the goal of disentangling the impact of each variable on the government's probability of success. Such a model with nine explanatory variables (not reported) enables us to establish that the significance of the variable DLAC stems almost exclusively from the significance of the variable DLG.[10] Recall that Model 2 established that (everything else equal) the government is more likely to lose when the defendant is the local government, the autonomous government, or the councils jointly considered. The results of our first refinement of Model 2 enable us to establish that such a result follows almost exclusively from the fact that the government is more likely to lose when the defendant is the local government. As a consequence of these results, we re-estimated Model 2 by replacing the variable DLAC by the variable DLG, thus obtaining *Model 3*. The results obtained for this model are summarized in Table A4, in part 2 of the appendix.[11]

Our second extension (now beginning from Model 3) has the purpose of determining whether there exists any relationship between the government's probability of success and the courts in which the cases are litigated. In order to perform this test, we sequentially introduce into Model 3 the variables CMDRD, CBRCLN, CBRGS, CVLLDLD, CLCRN, CVLNC, CGRND, CMLG, CSVLL, CLPGC, and CSCT in order to model the impact of the courts in the cities of Madrid, Barcelona, Burgos, Valladolid, La Coruña, Valencia, Granada, Málaga, Seville, Las Palmas (Gran Canaria) and Santa Cruz (Tenerife), respectively. We follow a sequential process in the sense that we estimated Model 3 eleven times, incorporating one court variable in each estimation. In other words, we first estimate Model 3 by incorporating the variable CMDRD; we then re-estimate Model 3 after dropping the variable CMDRD and adding the variable CBRCLN, and so forth, thus incorporating, one at a time, all the court variables.

---

[10] The disaggregation of the variable DLAC caused a problem in the estimation of the model; both the estimated coefficient and its standard error of the variable DAG were far larger that those estimated for the rest of the variables. This was due to the fact that autonomous governments were defendants in only six cases (out of 1,208), and in the six cases the government won. Thus, the combination of very few observations and identical results caused the estimation problem referred to.

[11] A comparison of Models 2 and 3 shows that the latter reaches a higher value in the likelihood function, has a slightly higher Cragg-Uhler R^2, and achieves the same in-sample prediction success.

We omit results for all these estimations but we summarize our main findings. [12] The estimation of the eleven models enables us to determine that the only courts that have a significant impact on the government's probability of success are those in La Coruña, Valencia, and Seville. More precisely, the government is (everything else equal) significantly more likely to win in Valencia, and significantly more likely to lose in La Coruña and Seville.

Based on the results of the eleven estimations, we re-estimated Model 3 including the variables CLCRN, CVLNC, and CSVLL all at once, thus obtaining *Model 4*; relevant results for this model are summarized in Table A5, in part 2 of the appendix. Note that this table shows that, when the three variables are simultaneously included in the model, the variable CSVLL becomes nonsignificant. Note, further, that a comparison of Models 3 and 4 establishes that the latter reaches a significantly-higher value in the likelihood function,[13] has a slightly higher Cragg-Uhler R^2, and achieves a slightly lower in-sample prediction success.

The last refinement we make is that of re-estimating Model 4 after eliminating the variable CSVLL, thus obtaining *Model 5*; relevant results for this model are summarized in Table A6, in part 2 of the appendix. A comparison of models 4 and 5 shows that the latter reaches a value not significantly lower in the likelihood function,[14] has a slightly lower Cragg-Uhler R^2, and achieves a slightly lower in-sample prediction success.

6.- Out-of-Sample Forecasting

As argued above, it is not very clear whether, in models with qualitative dependent variables, goodness-of-fit statistics can be interpreted as the proportion of the variability in the dependent variable explained by the independent variables. Thus, one of the usual ways to assess the explanatory power of these models is to evaluate their out-of-sample predictive ability. We thus evaluate the out-of-sample forecasting performance of two models: one that contains all the variables that entered the analysis in its different stages, and Model 5.

---

[12] During the estimation of these eleven models, the variable CLPGC caused a problem similar to that caused by the variable DAG before. That is, both the estimated coefficient and its standard deviation were far larger that those estimated for other variables. This problem stems from the fact that only six cases (out of 1,208) were tried in Las Palmas, and the government won all six. Thus, the combination of very few observations and identical results caused the estimation problem referred to.

[13] LRT = 2*(788.44-780.45) = 15.98 > $\chi^2_{3,.05}$ = 7.81.

[14] LRT = 2*(781.25-780.45) = 1.66 > $\chi^2_{1,.05}$ = 3.84.

During the estimation of the first of these two models (a model with twenty nine explanatory variables) we encountered again the estimation problems referred to above with the variables DAG and CLPGC. We thus performed the out-of-sample forecasting from a model with all the explanatory variables except for DAG and CLPGC; relevant results for this model (*Model 6*) are reported in Table A7, in part 2 of the appendix.[15] The out-of-sample forecasting performance of this model for 100, 200, and 300 observations is summarized below in Table 1; the out-of-sample forecasting performance of Model 5, on the other hand, is summarized in Table 2.

Table 1: Out-of-Sample Forecasting (Model 6)

100 Sentences				200 Sentences				300 Sentences			
		Actual				Actual				Actual	
		0	1			0	1			0	1
Predicted	0	38	24	Predicted	0	68	54	Predicted	0	103	77
	1	15	23		1	35	43		1	59	61
Correct Predictions:			61	Correct Predictions:			111	Correct Predictions:			164
% Correct Predictions:			61	% Correct Predictions:			55.5	% Correct Predictions:			54.7
Mean Square Error:			.24	Mean Square Error:			.25	Mean Square Error:			.25

Table 2: Out-of-Sample Forecasting (Model 5)

100 Sentences				200 Sentences				300 Sentences			
		Actual				Actual				Actual	
		0	1			0	1			0	1
Predicted	0	36	23	Predicted	0	77	57	Predicted	0	98	65
	1	17	24		1	26	40		1	64	73
Correct Predictions:			60	Correct Predictions:			117	Correct Predictions:			171
% Correct Predictions:			60	% Correct Predictions:			58.5	% Correct Predictions:			57
Mean Square Error:			.23	Mean Square Error:			.24	Mean Square Error:			.23

Table 1 shows that Model 6 achieves an out-of-sample forecasting success between 55% and 61%; Table 2, on the other hand, shows that Model 5 achieves an out-of-sample prediction success between 57% and 60%. A comparison of both tables shows that the predictions from the much simpler Model 5 are slightly better than those from Model 6. This can be gathered not only from the percentage of correct predictions of each model, but also from the mean square errors of each set of predictions. In short, these results suggest

[15] The variable CMDRD is not in the table because cases tried in Madrid constitute the base category of the court variables. Further, the variable DLAC is not in the table because it was disaggregated into its three components (DLG, DAG, and DC); recall, however, that the variable DAG was finally dropped due to the estimation problems mentioned above.

that the simpler model is slightly better than the more complicated one as a tool for predicting the results of the tax-related cases litigated by the government.[16]

7.- Computation of Probabilities

The next step in our empirical investigation is that of estimating the sensitivity of the government's probability of success with respect to changes in the explanatory variables included in the model. As is well known, in probit models the sensitivity of the dependent variable with respect to changes in any given independent variable cannot be interpreted directly from the estimated coefficients. This follows from the fact that these coefficients measure the impact of each independent variable on an index (M_i^* in our model) rather than on the dependent variable. In other words, the estimated relationship is linear between the explanatory variables and the index, but not between the explanatory variables and the dependent variable of interest. This last relationship actually depends on the specific point in which it is evaluated; we thus follow standard practice and take as a reference point the vector of means of the independent variables.[17]

We compute the sensitivity of the government's probability of success with respect to changes in the explanatory variables from Models 5 and 6. The results estimated for the latter model are reported immediately below in Table 3.

Table 3: Probabilities (Model 6)

Variable	Probability	Variable	Probability	Variable	Probability
TCG	10.0759	PI	-14.3832	CBRCLN	1.5204
TSF	-4.1401	DLG	-28.6715	CBRGS	11.0819
TI	-29.2621	DC	-38.6917	CVLLDLD	8.6151
TCT	0.6834	ICRTC	-18.3174	CLCRN	-8.9291
TRE	21.8816	ICCTC	-51.6603	CVLNC	-12.3006
SS350	0.6794	LFS	0.0006	CGRND	-7.8971
S3501M	-3.9689	RA	-14.8136	CMLG	-2.3078
SL1M	-4.0410	RLA	-10.6300	CSVLL	8.1401
PPF	-17.9500	RLAA	-11.9796	CSCT	3.6182

[16] Note from Tables 1 and 2 that the out-of-sample forecasts of Models 5 and 6 are "pessimistic." This follows from the fact that, as also noted above, in most of the incorrect predictions the error is that the model predicts that the government will lose when in fact the government won.

[17] Technically, the probabilistic impact on a dependent variable (y) that follows from a change in any given independent variable (x_i) is given by $\partial y / \partial x_i = \phi(\bar{x}\hat{\beta})\hat{\beta}_i$, where $\bar{x}$ is the vector of means of the independent variables, $\hat{\beta}$ is the vector of estimated coefficients, and ϕ is the standard Normal probability distribution.

Each number in the preceding table shows (everything else equal) the change in the government's probability of success when the case concerns a given variable. Or, put differently, it shows the change in the government's probability of success when each variable takes a value of 1 (as opposed to 0) and all the other variables are held constant at their mean values. For example, everything else equal, the government's probability of success increases by 10.08% when the case concerns capital gains, decreases 17.95% when the plaintiff is a private firm, and decreases 12.30% when the case is tried in the city of Valencia.

Given that not all the variables incorporated in Model 6 are significant, we re-estimated the sensitivity of the government's probability of success due to changes in the explanatory variables from Model 5. Such results are reported below in Table 4.

Table 4: Probabilities (Model 5)

Variable	Probability	Variable	Probability	Variable	Probability
TCG	9.1465	PPF	-18.1149	ICCTC	-35.5362
TI	-29.0740	PI	-14.2998	CLCRN	-13.0191
TRE	19.6358	DLG	-11.4767	CVLNC	-13.9391

Note that the sensitivities estimated with respect to variables included in both Model 5 and Model 6 are similar, except in the case of the variables DLG and ICCTC. Note, further, that some of the variables included in Model 6 (but excluded from Model 5) have a larger probabilistic impact on the government's probability of success that some of the variables that survived the sequential estimation.[18] With respect to this point, it should be kept in mind that the probabilities reported in Tables 3 and 4 are computed by taking into account the estimated coefficients but not their standard errors.

8.- Use of the Estimated Models

The next two examples briefly illustrate how the models could be used by the government to forecast its probability of success (and, perhaps, act accordingly) in two hypothetical cases. In both examples, the probabilities are estimated from both Model 5 and Model 6.

Example 1: The government must decide whether to try a case concerning the real estate tax, with an amount at stake of 200,000 pesetas, in which an individual is represented by

[18] The three variables representing the amount of resources spent by plaintiffs in Model 6, for example, affect the government's probability of success by more than 10% each (in absolute value), whereas the variable representing the capital gains tax, which survived the iterative process, affects such probability by only 9%.

a legal administrator. The case will go through a Regional Tax Court, it is expected to last 900 days, and will be tried in the city of Seville. Given this information, the government's probability of success is 72.41% according to Model 5,[19] and 80.38% according to Model 6.[20] In other words, the government should try this case.

Example 2: The government must decide whether to try a case concerning the income tax, with an amount at stake of 2 million pesetas, in which an individual is represented by both a legal administrator and an attorney. The case will go through the Central Tax Court, it is expected to last 100 days, and will be tried in the city of Valencia. Given this information, the government's probability of success is 3.02% according to Model 5,[21] and 2.57% according to Model 6.[22] In other words, the government should not try this case.

IV- CONCLUSIONS

Our main purpose has been to identify the factors that determine the government's probability of success in tax-related cases, and to use those factors to forecast such probability. We have also tried to make a contribution to the theory of litigation by putting our results within the context of such theory, and to illustrate some interesting characteristics of the Spanish litigation process. As argued above, we believe that our data had not been analyzed before with the approach we used in this article.

We started out by emphasizing an important difference among the cases in our sample, namely, the fact that in some types of conflicts settlements are possible case and in some other types they are not. We thus outlined two simple models of litigation (one for each type of case), and argued that the existence or absence of the settlement option is an important factor determining the government's frequency of success at trial. In particular we established that, when settlement is possible, "bad" cases (from the plaintiff's point of view) tend to be settled out of court, thus decreasing the probability that a "bad" case reaches a Court of Appeals. We also discussed other variables that may affect, directly or indirectly, the government's

[19] $\Phi(0.46066+0.49441-0.36006) = \Phi(0.5950) = 0.7241$.

[20] $\Phi(1.1644+0.55105+0.01711-0.36222-0.46129+0.00001*900-0.2677+0.20499) = \Phi(0.8553) = 0.8038$.

[21] $\Phi(0.46066-0.73206-0.36006-0.89477-0.35097) = \Phi(-1.8772) = 0.0302$.

[22] $\Phi(1.1644-0.73692-0.10177-0.36222-1.301+0.00001*100-0.30169-0.30977) = \Phi(-1.9478) = 0.0257$.

probability of success, like the existence of intermediate courts, risk aversion, budget constraints, and variables determining threshold probabilities.

We argued that, due to the particular characteristics of the area of the law that generated the cases in our sample, there is no compelling reason to expect any particular frequency of the government's success at trial, like, for example, the 50% rate predicted by Priest and Klein (1984). However, we also argued that there exist factors that tend to center the range of outcomes, like the government's budget constraint and the plaintiff's thresholds and risk aversion.

Our basic hypothesis in the empirical analysis was that the government's probability of success was determined by the type of taxes involved in the litigation, the stakes of the cases, the identity of the plaintiffs and the defendants, the existence of intermediate courts, the length of the litigation process, the amount of resources spent by plaintiffs, and the court in which the cases were tried. Given the theoretically ambiguous impact of many of these variables on the government's probability of success, we hope to have produced results that give a step forward in the direction of resolving some of these ambiguities.

We have found that the government's probability of success is negatively correlated with cases concerning the income tax, cases in which the plaintiff is a private firm or an individual, cases in which the defendant is the local government, and cases that have gone through the filter of the Central Tax Court. We have found, on the other hand, a positive correlation between the government's probability of success and cases concerning capital gains taxes and real estate taxes. Finally, we have found that the government is more likely to lose in Valencia and La Coruña, and more likely to win in Seville. All in all, we believe that our results, quite in line with the main propositions of the theory of litigation, may be a useful empirical starting point to address the open questions left by our current knowledge of the litigation process.

APPENDIX

1- NOTATION AND DESCRIPTIVE ANALYSIS

SNTNC: Sentence (=1 if the government wins the case; =0 otherwise)

TCG: The case is about the capital gains tax.
TSF: The case is about the service fees tax.
TI: The case is about the income tax.
TCT: The case is about the capital transactions tax.
TRE: The case is about the real estate tax.

SS350: The stakes of the case are larger than 350,000 pesetas.
S3501M: The stakes of the case are between 350,000 and 1 million pesetas.
SL1M: The stakes of the case are larger than 1 million pesetas.

PPF: The plaintiff is a private firm.
PI: The plaintiff is an individual.

DLAC: The defendant is a Local government, an Autonomous government, or a Council.
DLG: The defendant is a Local government.
DAG: The defendant is an Autonomous government.
DC: The defendant is a Council.

ICRTC: The case goes through a Regional Tax Court.
ICCTC: The case goes through the Central Tax Court.

LFS: Length (in days) between the filing of the case and the sentence.

RLA: The plaintiff is represented by a legal administrator.
RA: The plaintiff is represented by an attorney.
RLAA: The plaintiff is represented by both a legal administrator and an attorney.

CMDRD: The case is tried in Madrid.
CBRCLN: The case is tried in Barcelona.
CBRGS: The case is tried in Burgos.
CVLLDLD: The case is tried in Valladolid.
CLCRN: The case is tried in La Coruña.
CVLNC: The case is tried in Valencia.
CGRND: The case is tried in Granada.
CMLG: The case is tried in Málaga.
CSVLL: The case is tried in Seville.
CLPGC: The case is tried in Las Palmas (Gran Canaria).
CSCT: The case is tried in Santa Cruz (Tenerife).

Table A1: Descriptive Statistics

Variable	Cases	Won	Lost	Mean	SD
SNTNC	1,208	561	647		
TCG	363	199	164	0.30050	0.45866
TSF	108	42	66	0.08940	0.28544
TI	120	28	92	0.09934	0.29924
TCT	75	43	32	0.06209	0.24141
TRE	47	30	17	0.03891	0.19345
SS350	435	218	217	0.36010	0.48023
S3501M	246	104	142	0.20364	0.40287
SL1M	277	121	156	0.22930	0.42056
PPF	509	220	289	0.42136	0.49398
PI	438	179	259	0.36258	0.48094
DLG	595	263	332	0.49255	0.50015
DAG	6	6	0	0.00497	0.07033
DC	12	4	8	0.00993	0.09921
ICRTC	534	268	266	0.44205	0.49684
ICCTC	36	8	28	0.02980	0.17011
LFS	1,208	----	----	1059.80000	614.36000
RA	286	120	166	0.23675	0.42527
RLA	339	160	179	0.28063	0.44949
RLAA	563	269	294	0.46606	0.49905
CMDRD	427	217	210	0.35348	0.47825
CBRCLN	186	78	108	0.15397	0.36107
CBRGS	39	21	18	0.03229	0.17683
CVLLDLD	39	18	21	0.03229	0.17683
CLCRN	101	37	64	0.08361	0.27692
CVLNC	179	65	114	0.14818	0.35542
CGRND	33	14	19	0.02732	0.16308
CMLG	24	10	14	0.01987	0.13960
CSVLL	156	86	70	0.12914	0.33549
CLPGC	6	6	0	0.00497	0.07033
CSCT	18	9	9	0.01490	0.12121

Means and standard deviations (SD) computed on the basis of 1,208 observations. “Won” and “Lost” both considered from the government’s point of view.

2- ECONOMETRIC ANALYSIS

Table A2: Model 1

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.23921	0.09754	2.4525	0.0618
TSF	-0.10182	0.14025	-0.7260	-0.0078
TI	-0.65910	0.15573	-4.2323	-0.0563
TCT	0.13215	0.17269	0.7652	0.0071
TRE	0.48583	0.20072	2.4204	0.0162
SS350	0.01173	0.10448	0.1123	0.0036
S3501M	-0.12138	0.11645	-1.0424	-0.0213
SL1M	-0.13117	0.11328	-1.1579	-0.0259
PPF	-0.41549	0.10655	-3.8994	-0.1505
PI	-0.40480	0.10800	-3.7483	-0.1262
DLAC	-0.38668	0.26751	-1.4455	-0.1690
ICRTC	-0.14991	0.26533	-0.5650	-0.0570
ICCTC	-0.93143	0.35781	-2.6031	-0.0239
LFS	0.00008	0.00006	1.2687	0.0740
RA	-0.40854	0.29914	-1.3657	-0.0832
RLA	-0.34298	0.29822	-1.1501	-0.0828
RLAA	-0.23643	0.29492	-0.8017	-0.0947
Constant	0.76005	0.41471	1.8327	0.6535
LLFM = -783.88	LLFC = -834.26	LRT = 100.74	CRAGG-UHLER R^2 = .1069	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 739
SUCCESS	PREDICTED	0	455	277	% Correct Predictions: 61.18
TABLE:		1	192	284	

EAM=Elasticity at Means; LLFM=log-likelihood function for the model; LLFC=log-likelihood function for the constant; LRT=likelihood-ratio test. Observations=1,208.

Table A3: Model 2

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.27761	0.08937	3.1064	0.0717
TI	-0.69446	0.14898	-4.6615	-0.0593
TRE	0.48284	0.19678	2.4537	0.0161
PPF	-0.41906	0.10189	-4.1130	-0.1517
PI	-0.38806	0.10556	-3.6762	-0.1209
DLAC	-0.29972	0.08412	-3.5630	-0.1308
ICCTC	-0.80528	0.24729	-3.2564	-0.0206
Constant	0.36608	0.09398	3.8954	0.3144
LLFM = -789.08	LLFC = -834.26	LRT = 90.36	CRAGG-UHLER R^2 = .0963	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 742
SUCCESS	PREDICTED	0	547	366	% Correct Predictions: 61.42
TABLE:		1	100	195	

Table A4: Model 3

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.28760	0.08986	3.2006	0.0742
TI	-0.69738	0.14867	-4.6907	-0.0595
TRE	0.48121	0.19651	2.4488	0.0161
PPF	-0.41791	0.10184	-4.1034	-0.1512
PI	-0.38472	0.10559	-3.6435	-0.1198
DLG	-0.31466	0.08421	-3.7367	-0.1331
ICCTC	-0.80472	0.24711	-3.2566	-0.0206
Constant	0.36438	0.09359	3.8935	0.3130
LLFM = -788.44	LLFC = -834.26	LRT = 91.64	CRAGG-UHLER R^2 = .0976	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 742
SUCCESS	PREDICTED	0	546	365	% Correct Predictions: 61.42
TABLE:		1	101	196	

Table A5: Model 4

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.25331	0.09331	2.7147	0.0654
TI	-0.71358	0.15026	-4.7489	-0.0609
TRE	0.51403	0.19805	2.5955	0.0172
PPF	-0.44457	0.10331	-4.3034	-0.1609
PI	-0.35033	0.10660	-3.2866	-0.1091
DLG	-0.29672	0.08508	-3.4877	-0.1256
ICCTC	-0.87514	0.24878	-3.5177	-0.0224
CVLNC	-0.32091	0.11119	-2.8861	-0.0408
CSVLL	0.14915	0.11569	1.2892	0.0165
CLCRN	-0.30045	0.13890	-2.1631	-0.0216
Constant	0.42019	0.10255	4.0973	0.3610
LLFM = -780.45	LLFC = -834.26	LRT = 107.62	CRAGG-UHLER R^2 = .1138	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 740
SUCCESS	PREDICTED	0	463	284	% Correct Predictions: 61.26
TABLE:		1	184	277	

Table A6: Model 5

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.23030	0.09164	2.5131	0.0595
TI	-0.73206	0.14943	-4.8991	-0.0625
TRE	0.49441	0.19737	2.5051	0.0165
PPF	-0.45612	0.10279	-4.4375	-0.1651
PI	-0.36006	0.10624	-3.3892	-0.1122
DLG	-0.28897	0.08484	-3.4063	-0.1223
ICCTC	-0.89477	0.24860	-3.5993	-0.0229
CVLNC	-0.35097	0.10874	-3.2276	-0.0447
CLCRN	-0.32781	0.13731	-2.3874	-0.0235
Constant	0.46066	0.09752	4.7238	0.3958
LLFM = -781.28	LLFC = -834.26	LRT = 105.95	CRAGG-UHLER R^2 = .1122	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 712
SUCCESS	PREDICTED	0	393	242	% Correct Predictions: 58.94
TABLE:		1	254	319	

Table A7: Model 6

Variable	Coefficient	Std.Error	t-ratio	EAM
TCG	0.25375	0.10186	2.4912	0.0656
TSF	-0.10426	0.14406	-0.7238	-0.0080
TI	-0.73692	0.16127	-4.5696	-0.0630
TCT	0.01721	0.18022	0.0955	0.0009
TRE	0.55105	0.20303	2.7141	0.0184
SS350	0.01711	0.11155	0.1534	0.0053
S3501M	-0.09995	0.12217	-0.8181	-0.0175
SL1M	-0.10177	0.11890	-0.8559	-0.0201
PPF	-0.45204	0.10911	-4.1428	-0.1639
PI	-0.36222	0.10923	-3.3160	-0.1130
DLG	-0.72204	0.24401	-2.9590	-0.3060
DC	-0.97439	0.45515	-2.1408	-0.0083
ICRTC	-0.46129	0.24217	-1.9048	-0.1754
ICCTC	-1.30100	0.35414	-3.6736	-0.0334
LFS	0.00001	0.00007	0.2072	0.0136
RA	-0.37306	0.30330	-1.2300	-0.0760
RLA	-0.26770	0.30236	-0.8854	-0.0646
RLAA	-0.30169	0.30022	-1.0049	-0.1210
CBRCLN	0.03829	0.12988	0.2948	0.0051
CBRGS	0.27908	0.23503	1.1874	0.0078
CVLLDLD	0.21696	0.22856	0.9492	0.0060
CLCRN	-0.22486	0.16214	-1.3869	-0.0162
CVLNC	-0.30977	0.14220	-2.1785	-0.0395
CGRND	-0.19888	0.24276	-0.8192	-0.0047
CMLG	-0.05812	0.27885	-0.2084	-0.0010
CSVLL	0.20499	0.14365	1.4270	0.0228
CSCT	0.09112	0.31030	0.2936	0.0012
Constant	1.16440	0.42277	2.7541	1.0017
LLFM = -773.62	LLFC = -834.26	LRT = 121.28	CRAGG-UHLER R^2 = .1276	

IN-SAMPLE			ACTUAL		
PREDICTION			0	1	Correct Predictions: 742
SUCCESS	PREDICTED	0	452	271	% Correct Predictions: 61.42
TABLE:		1	195	290	

REFERENCES

Eisenberg, Theodore (1990). "Testing the Selection Effect: A New Theoretical Framework with Empirical Tests." Cornell Law Review, 77, 1124.

Galanter, Mark (1974). "Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change." Law and Society Review, 9, 95.

Goldberger, A.S. (1973). "Correlations Between Binary Choices and Probabilistic Predictions." Journal of the American Statistical Association, 68, 84.

Hylton, Keith (1993). "Asymmetric Information and the Selection of Disputes for Litigation." Journal of Legal Studies, 22, 187.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

Morrison, A.G. (1972). "Upper Bounds for Correlations Between Binary Outcomes and Probabilistic Predictions." Journal of the American Statistical Association., 68, 7.

Pastor, Santos (1993). "Composición, Estructura y Características de la Litigiosidad Contencioso-Administrativa: Un Análisis Muestral." Boletín de Información del Consejo General del Poder Judicial, 111, 205.

Pastor, Santos (1994). "Qué Hacer con los Conflictos Fiscales?" Hacienda Pública Española, 132, 225.

Perloff, Jeffrey, Daniel Rubinfeld, and Paul Ruud (1993). "Antitrust Settlements and Trial Outcomes." Unpublished manuscript.

Priest, George, and Benjamin Klein (1984). "The Selection of Disputes for Litigation." Journal of Legal Studies, 13, 1.

Shavell, Steven (1985). "Any Frequency of Plaintiff Victory at Trial Is Possible." Discussion Paper 7/95, Program in Law and Economics, Harvard University.

Wittman, Donald (1985). "Is the Selection of Cases for Trial Biased?" Journal of Legal Studies, 14, 185.

Wittman, Donald (1988). "Dispute Resolution, Bargaining, and the Selection of Cases for Trial: A Study of the Generation of Biased and Unbiased Data." Journal of Legal Studies, 17, 313.